

Radiomics for Coronary Plaque Analysis from Cardiac Computed Tomography Angiography

Radiomics zur Plaqueanalyse in den Herzkranzgefäßen mittels Computertomographie des Herzens

Master's Thesis in Medical Image and Data Processing

submitted
by

Felix Denzinger

born 30.10.1991 in Neustadt a. d. Waldnaab

Written at

Lehrstuhl für Mustererkennung (Informatik 5)
Department Informatik
Friedrich-Alexander-Universität Erlangen-Nürnberg.

Advisor: Dr.-Ing. Michael Wels, Nishant Ravikumar Ph.D.,
PD Dr.-Ing. Tino Haderlein, Prof. Dr.-Ing. habil. Andreas Maier

Started: 07.06.2018

Finished: 07.12.2018

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Die Richtlinien des Lehrstuhls für Studien- und Diplomarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts.

Erlangen, den 06.12.2018

Übersicht

Die Erkennung und Klassifizierung von risikobehafteten Plaquesegmenten in den Herzkranzgefäßen ist für verbesserte klinische Ergebnisse von hoher Bedeutung. Daher versuchen wir in dieser Arbeit mittels zweier Ansätze — basierend auf Radiomics und Deep Learning — Plaquesegmente bezüglich ihres Potentials kardiale Ereignisse auszulösen zu analysieren. Dies kann durch Identifizierung folgender Eigenschaften von Hochrisikoplaques mittels Verwendung von CCTA-Scans beurteilt werden: niedrige Hounsfield-Unit-Werte, positives Remodeling, punktuelle Kalzifikationen und ein hoher Stenosegrad. Um diese Ansätze zu evaluieren wurde eine Datensammlung mit 463 Läsionen von 105 Patienten herangezogen. Der Begriff Radiomics beschreibt für den Menschen teilweise nicht wahrnehmbare Bildinformationen. Diese Bildmerkmale basieren auf Statistiken erster und höherer Ordnung, sowie formbasierten Merkmalen, welche unter verschiedenen Bildtransformationen berechnet werden. Sie können mittels Mustererkennungsalgorithmen zur Klassifikation der Plaquesegmente verwendet werden. Aus den daraus resultierenden Merkmalsvektoren wurden jene Merkmale mit der höchsten Unterscheidungskraft verwendet, um unsere Zielgrößen mittels unterschiedlicher Klassifizierer vorherzusagen. Für unseren Deep Learning Ansatz wurde für jeden Plaque ein 3D-Bild orthogonal zur Gefäßmittellinie interpoliert. Dieses 3D-Bild wurde anschließend in eine Sequenz von sich überlappenden Würfeln zerschnitten und mit einem neuralen Merkmalsextraktions- und Sequenzanalysenetzwerk verarbeitet um unsere Zielgrößen vorherzusagen. Beide Ansätze wurden mit Hilfe einer zehnfachen Kreuzvalidierung unter verschiedenen Hyperparametern evaluiert. Unsere Ansätze eigneten sich sehr gut für die Identifizierung von hohem Stenosegrad (AUC: 0.91; MCC: 0.59), niedrigen HU-Werten (AUC: 0.96; MCC: 0.65) und punktuellen Kalzifikationen (AUC: 0.86, MCC 0.69). Darüber hinaus konnten unsere Algorithmen zuverlässig Revaskularisierungsentscheidungen auf unterschiedlichen Ebenen — namentlich Segment, Hauptast und Patient — vorhersagen (AUC: 0.82, 0.75, 0.79; MCC: 0.47, 0.39, 0.41). Allerdings konnten sie das positive Remodeling nur schlecht prognostizieren (AUC: 0.69; MCC: 0.26). Die Evaluierungsmaßzahlen bezüglich der Hauptzielgröße — der Revaskularisierungsentscheidung auf Segmentebene — liegen im Bereich anderer State of the Art Methoden, welche eine AUC, einen positiven Vorhersagewert, negativen Vorhersagewert, Sensitivität und Spezifität zwischen 0.79-0.93, 0.56-0.74, 0.84-0.95, 0.76-0.89 und 0.61-0.95 haben, da unser Deep Learning Ansatz 0.82, 0.53, 0.91, 0.67 und 0.85 und unser Radiomics Ansatz 0.81, 0.62, 0.87, 0.44 und 0.93 erzielen. Diese Arbeit zeigt also, dass hochriskante Plaques der Koronararterien automatisch aus CCTA-Scans mit den genannten Ansätzen identifiziert werden können.

Abstract

The identification and analysis of risk inducing coronary artery plaque segments is an important task to improve patient management and clinical outcomes. In this thesis, we use two machine learning approaches based on radiomics and deep learning in order to analyse coronary artery plaques with respect to their probability to cause adverse cardiac events. This can be assessed by identifying different high risk plaque characteristics from coronary computed tomography angiography (CCTA) scans: low Hounsfield unit attenuation, spotty calcifications, positive remodeling and a high stenosis grade. For this task a data collection containing 463 lesions from 105 patients was used to evaluate these machine learning algorithms. Radiomics describes extracted image information, which is partly not perceivable to the human reader. These image features can be processed with standard machine learning algorithms and include first- and higher order statistics and shape-based features calculated under different transformations. From the resulting feature vectors the features with the highest discriminative power were used in order to predict our targets with several different classifiers. For the deep learning approach a multi planar reformatted image stack was extracted from the CCTA scan by interpolating images orthogonal to the vessel centerline. This stack was cut into a sequence of overlapping cubes and then processed by a feature extraction and sequence analysis neural network to gain a final prediction. Both approaches were evaluated using ten-fold cross-validation and different hyper parameters. We reached very good scores for the targets high stenosis degree (area under curve (AUC): 0.91; Matthews correlation coefficient (MCC): 0.59), low HU values (AUC: 0.96; MCC: 0.65) and spotty calcifications (AUC: 0.86, MCC 0.69). Furthermore, our methods performed well in terms of predicting the need for revascularisation on different levels, namely, segment, branch and patient (AUC: 0.82, 0.75, 0.79; MCC: 0.47, 0.39, 0.41). However, they performed poorly in terms of predicting positive remodeling (AUC: 0.69; MCC: 0.26). For most metrics our main target — the revascularisation need on lesion level — performed comparably to state of the art methods which have an AUC, positive predictive value, negative predictive value, sensitivity and specificity scores between 0.79-0.93, 0.56-0.74, 0.84-0.95, 0.76-0.89 and 0.61-0.95, since our deep learning approach scored 0.82, 0.53, 0.91, 0.67 and 0.85 and our radiomics approach scored 0.81, 0.62, 0.87, 0.44 and 0.93. This thesis can therefore be seen as a proof of concept that high risk coronary artery plaques can be automatically identified from CCTA scans using the mentioned approaches.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Related Work	5
1.2.1	Stenosis Assessment	5
1.2.2	Plaque Assessment	7
1.3	Baseline	9
1.4	Contribution	9
1.5	Overview	10
2	Theory	11
2.1	Goals of the Thesis	11
2.2	Medical Background	12
2.2.1	Anatomy	12
2.2.2	Plaque Pathogenesis	12
2.2.3	Plaque Treatment	13
2.3	Radiomics	14
2.3.1	Radiomic Features	14
2.3.2	Feature Selection	20
2.3.3	Classifiers	21
2.4	Deep Learning	26
2.4.1	Constituents	26
2.4.2	Advanced Concepts	31
2.5	Evaluation Metrics	32
3	Material and Methods	35
3.1	Data	35

3.1.1	Data Analysis	36
3.1.2	Data Cleansing	38
3.2	Radiomics Based Approach	38
3.2.1	Preprocessing	38
3.2.2	Radiomics Classification	39
3.3	Neural Network Based Approach	41
3.3.1	Preprocessing	41
3.3.2	Deep Learning Classification	42
3.4	Prototype Integration	44
4	Results	45
4.1	Evaluation	45
4.1.1	Radiomics Based Approach	45
4.1.2	Neural Network Based Approach	48
4.1.3	Comparison between Deep Learning and Radiomics	50
4.2	Comparison to Related Work	51
5	Summary and Outlook	55
5.1	Discussion	55
5.2	Future Work	56
5.3	Summary	57
5.4	Conclusion	60
A	Data Analysis	65
B	Models	67
C	Detailed Results	73
C.1	Radiomics Mask Comparison	74
C.2	Selected Features	78
C.3	Radiomics Deep Learning Comparison	81
List of Figures		85
List of Tables		87
Bibliography		89

Chapter 1

Introduction

1.1 Motivation

Medical Background

Cardiovascular diseases (CVDs) have persisted to be the leading cause of death across all developed countries, over the last decade [Men15]. Most CVDs, like ischemia or cardiac infarction, are related to atherosclerotic plaques in the associated arteries [Nag03]. These plaque segments arise due to stress or inflammatory reactions between the inner and outer vessel wall. This leads to immune response cells like macrophages, T-cells or smooth muscle cells aggregating between the vessel walls and forming a necrotic core. The resulting atherosclerotic plaque might rupture or significantly narrow the lumen. Plaque rupture leads to thrombus formation and may therefore induce stroke or cardiac infarction. If the lumen is significantly narrowed due to the plaque, the heart muscle is not sufficiently perfused by blood inducing cardiac ischemia [SG14]. Two types of high risk plaque segments exist: functionally significant plaques, which immediately lead to cardiac ischemia and vulnerable plaques, which lead to adverse coronary syndromes (ACS) like stroke or cardiac infarction. A schematic with the characteristics of these high risk plaque segments is shown in Figure 1.1.

Gold Standard

The reference standard measure to judge whether a plaque segment is functionally significant and needs to be revascularised is the fractional flow reserve (FFR) value, which is defined as the intravascular pressure before and after the segment. As of today, the actual FFR measurement is still performed in an interventional setup, where a catheter is guided to the location of the

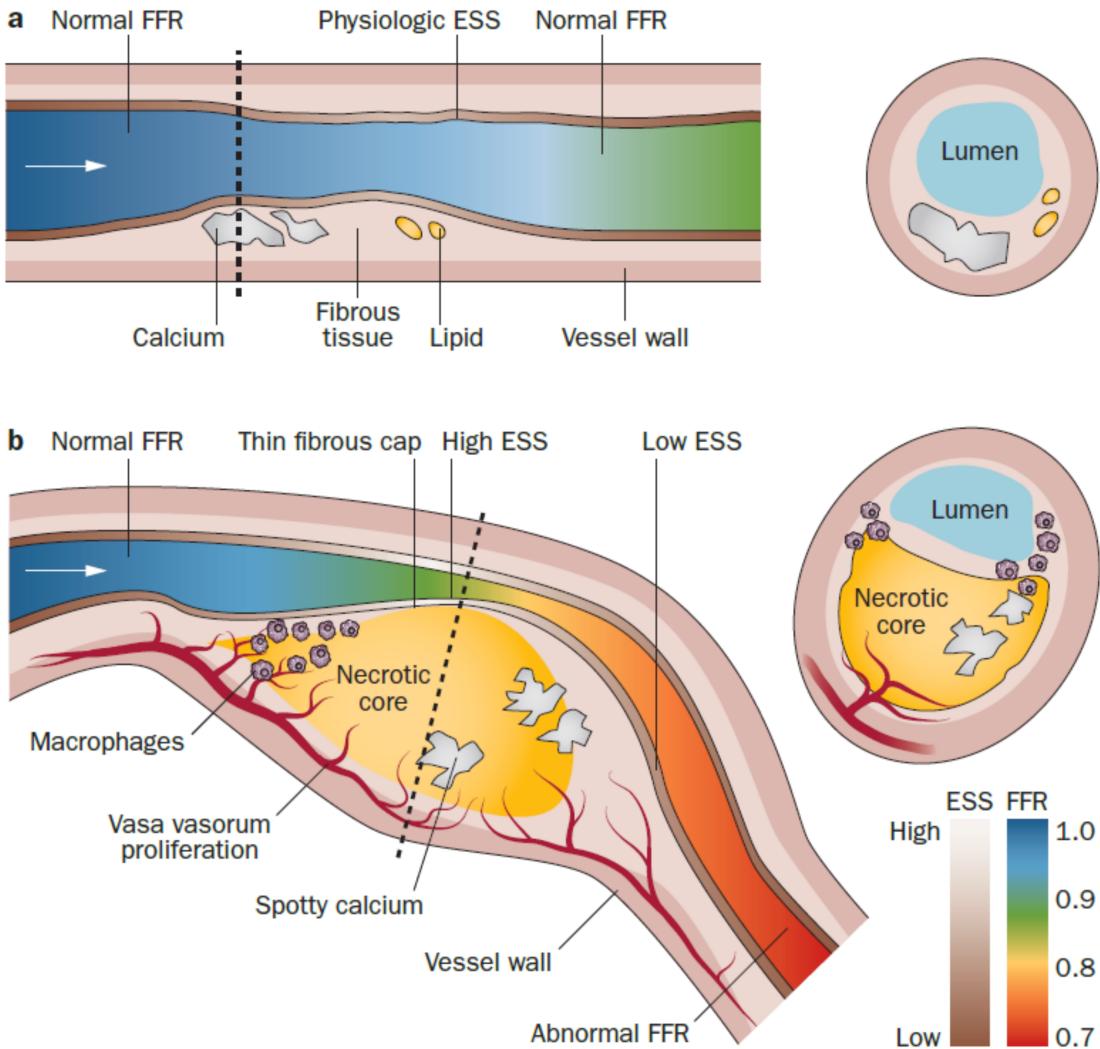


Figure 1.1: Morphological characteristics of stable (a) or vulnerable (b) plaques [MH14]. Vulnerable plaques originate from high endothelial shear stress (ESS) induced inflammatory reactions leading to a accumulation of immune cells (macrophages), which form a necrotic core together with smooth muscle cells and T-cells. The FFR value corresponds to the blood pressure before and after a stenotic segment.

stenosis and the FFR value is measured [Pij96]. An alternate approach for the stenosis assessment is simulated FFR [Tay13], which tries to simulate the FFR values from coronary computed tomography angiography (CCTA) data using a fluid dynamics approach, which needs a very exact segmentation and computational mesh generation, which can sometimes only be guaranteed through time consuming manual interaction. In order to assess the vulnerability of a coronary plaque segment intravascular ultrasound (IVUS) proved to be the best approach [Cal11]. However, IVUS is also performed in an interventional setup and determining plaque vulnerability actually

has little direct impact on revascularisation decisions [Cur16]. Thus, it is rarely applied in the clinical workflow.

Coronary Computed Tomography Angiography

Since interventions induce stress and risk for the patient, gaining information about the functional significance or vulnerability of a segment with non-invasive modalities is of high interest. One modality suitable for this task is CCTA. CCTA provides means to differentiate between plaque segments with distinct features. These include: a high stenosis degree, defined as the ratio of narrowing; positive remodeling, defined as an enlargement of the vessel into the perivascular space, caused by the plaque; low Hounsfield unit (HU) values; spotty calcification, which denote small calcified volumes inside the plaque and the napkin-ring sign as seen in Figure 1.2. Besides those features, which provide information about the vulnerability of a segment [MH14], the functional stenosis degree of a plaque segment can be assessed by comparing the cross-sectional area of a presumed healthy vessel with the area at the tightest narrowing [Sch06] [Cur16]. In order to automatically assess both functional significance and vulnerability of plaque segments this thesis analyses a data collection of CCTA scans of 105 patients with a total of 464 plaque segments. This is performed with two approaches: radiomic feature analysis and neural networks.

Radiomics

The term “radiomics” describes the extraction of large amounts of image features from radiographic data. These features include first or higher-order statistics and shape-based features under different image transformations [Lam12]. Originating from cancer analysis, recent research proved radiomics to be applicable for coronary plaque analysis as well [Kol18]. The high dimensional feature vectors do not only allow the distinction of measurement-based plaque characteristics like positive remodeling, low attenuation and spotty calcification, but also the distinction of qualitative characteristics like the napkin-ring sign (NRS) [Kol17]. In this thesis we use radiomic features for classification on multiple targets. On the one hand we want to predict the need for revascularisation on multiple abstraction levels (plaque, vessel and patient level) along the coronaries. And our second goal is the detection of vulnerable plaques. As labels for future ACS are close to impossible to create (see Section 1.2.2), we predict the presence of high risk plaque features as a surrogate for the actual target of interest. The radiomic feature calculation needs a segmentation mask of the area of interest, from which – in combination with the original image – a high dimensional (> 1200) feature vector is calculated. Since feature vectors of this size often contain redundant or irrelevant features, a feature selection method is used to reduce

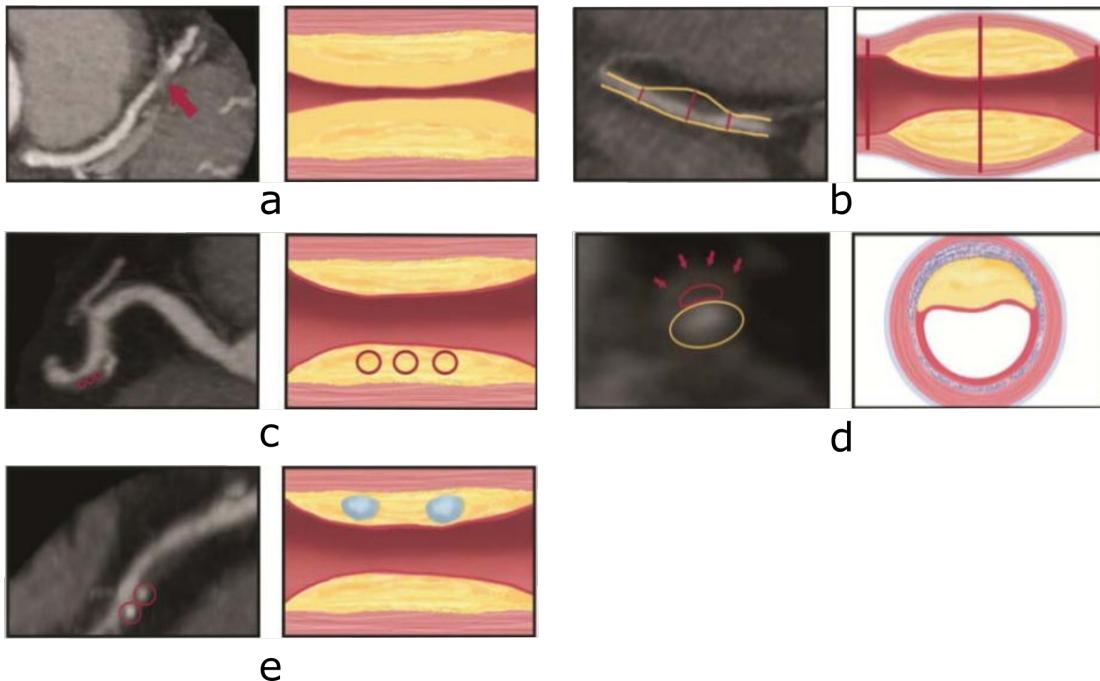


Figure 1.2: Different plaque features and their shape in CCTA [Puc14]. (a) Stenosis degree of $> 50\%$; (b) positive remodeling; (c) low HU attenuation; (d) napkin ring sign; (e) spotty calcifications

its dimensionality. The resulting feature vectors are then used to perform a classification. In the scope of this thesis four different classifiers — support vector machines (SVMs), gaussian naïve Bayes (GNB), random forests and extreme gradient boosting (XGBoost) — are evaluated with different masks — lumen, plaque, and lumen combined with plaque — and the associated selected features.

Neural Networks

Our alternative approach for the assessment of coronary artery plaque is based on deep learning. In [Zre18] the authors showed that analysis the coronary arteries using deep learning algorithms was feasible. For this approach multi planar reformatted (MPR) image stacks along the coronary artery centerline are extracted and cut into a sequence of overlapping image cubes. Then for each cube image features are extracted using different convolutional neural networks (CNNs) before performing a sequence analysis with either a recurrent neural network (RNN) or a fully convolutional approach. In the scope of this thesis four combinations of feature selection and sequence analysis networks are evaluated with respect to their capability to predict high risk plaque features and the need for revascularisation. Finally, both methods are integrated in a computer

assisted diagnosis (CAD) system with the target of predicting the need for revascularisation. The resulting CAD program is implemented in form of a coronary plaque analysis program.

1.2 Related Work

As discussed previously, the main topic of interest in this thesis is to differentiate between functional significance and vulnerability of plaque segments. In order to provide some context for the methods investigated, relevant literature is discussed in subsequent sections.

1.2.1 Stenosis Assessment

Reference standard for the classification of the functional severity of a stenosis with CCTA is visual assessment based on severity measurements. In order to standardize findings the coronary artery disease reporting and data system (CAD-RADS) was introduced in [Cur16]. This system categorizes stenoses into six groups based on the percentage of maximal narrowing in relation to the suggested healthy vessel: no (0%), minimal (1%-24%), mild(25%-49%), moderate(50%-69%), severe(70%-99%) stenosis and total occlusion for 100% narrowing. Depending on the category of the patient's most severe plaque segment further patient management is adapted. For categories up to mild, no further cardiac investigation is recommended and different preventive therapies are adopted, to prevent progression of stenosis severity. However, for all other categories functional assessment regarding cardiac ischemia, like for example invasive FFR measurement should be considered [Cur16]. Visual assessment comes with some major problems: pre-processing methods vary depending on the used system as does the resolution and therefore the measurement of the stenosis itself. Moreover, for approximating the volume of a hypothetically stenosis free vessel, the start and end point of the stenosis segment is averaged with respect to the position of the tightest narrowing. This makes the measurement ill-defined for segments in branches with an overall high plaque burden and segments at bifurcations, where the reference measurements do not enable an exact approximation (see Figure 1.3). Another drawback of the CAD-RADS scoring: only the highest degree stenosis is considered for the patient score even though there is strong evidence, that the overall plaque burden plays a very important role for CVDs [AZ15].

Semi-automatic approaches to classify the severity of a lesion are for example described by the authors of reference [Kir13]. Most of these approaches rely on exact lumen segmentation and further detection of plaque presence. However, a high amount of manual interaction is still needed which increases for arteries with high degree of calcification or movement artefacts [Kir13, Zre18]. Another recent approach for assessing the functional relevance of a plaque segment using CCTA is

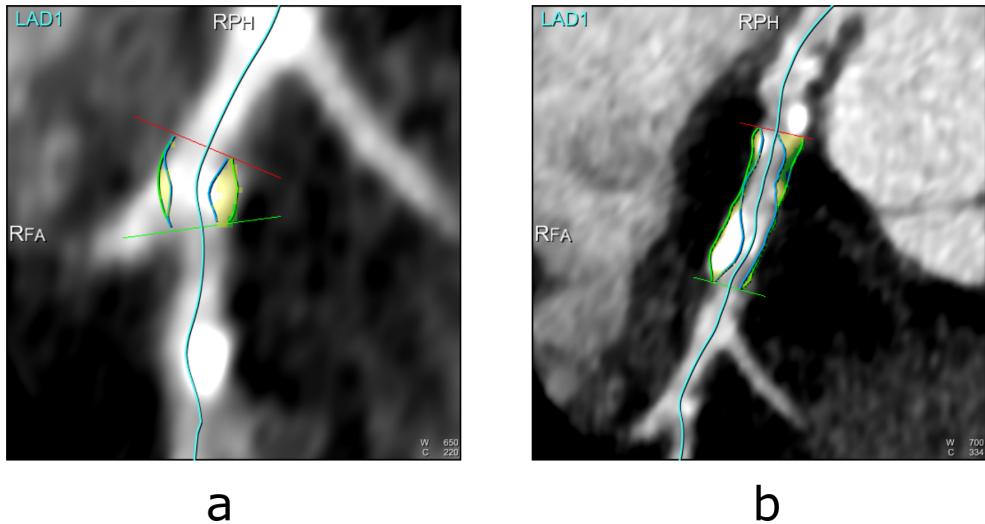


Figure 1.3: CPR view of ill defined plaque segments. (a) plaque at bifurcation: Reference measurements do not consider the natural vessel narrowing caused by the bifurcation; (b) plaque in branch with high plaque burden.

the simulated FFR measurement, proposed by [Tay13]. The workflow of this method is visualized and described in Figure 1.4. The core idea of this approach is to view the coronary artery system as a fluid dynamic system with a source (the aorta) and a sink (the left myocardium) and to evaluate partial differential equations based on the microvascular resistances of coronary artery branches. But Figure 1.4 also emphasises the massive amount of pre-processing steps necessary to gain the FFR information, most of them need manual interaction or confirmation and can be prone to error dependent on the dataset at hand. Especially exact segmentation is vulnerable to movement artefacts, inter-/intra-operator variation and good image quality. The overall time consumption for this method remains high, even with newest approaches providing massive speed up by using machine learning algorithms [Itu16] and fast interactive tools for segmentation [Wel16]. An approach similar to our method is shown in reference [Zre18]. It requires solely centerline extraction as pre-processing step. By interpolating MPR images orthogonal to every centerline point a straightened image stack is extracted for each vessel. In order to perform analysis of plaques with differing length and in order to incorporate spatially correlated image information, the image stack is then separated into a sequence of overlapping cubes. By combining the feature extraction capabilities of CNNs and the sequence analysis capabilities of RNNs plaque analysis with respect to the stenosis degree could be performed. For this only one cut off value of 50% narrowing was considered. This method not only allows assessment of the stenosis but also detection of relevant segments, but from our point of view lacks a proper cross-validation to verify

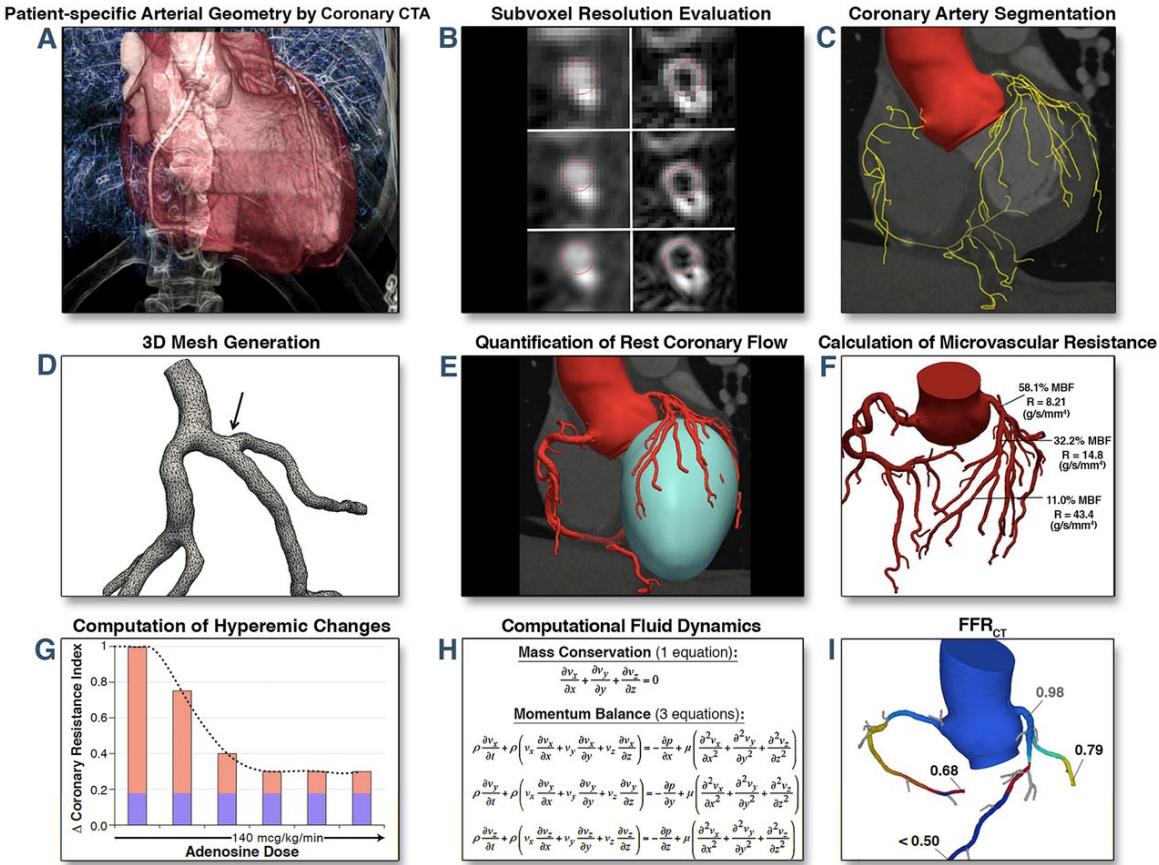


Figure 1.4: Workflow of noninvasive FFR [Min15]. (a) CCTA scan; (b) super scaling using subvoxel resolution evaluation techniques; c) centerline extraction and lumen segmentation of the second- and third order vessels; (d) creation of 3D meshes using tetrahedral vertices in order to enable continuity at branch points; (e) calculation of the rest coronary flow by quantifying the left ventricular mass as the sink; (f) microvascular resistance calculation using the vessel volume, flow and resistance; (g) considering the reduced coronary resistance due to the patients adenosine dose [Wil90]; (h) calculate partial differential equation based on the Navier-Stokes equations, which define the fluid dynamics blood flow; (i) visualisation of the FFR values for the segmented volume [Min15]

the results.

1.2.2 Plaque Assessment

In order to assess the plaque structure with respect to its probability to rupture current reference standard is IVUS [Cal11]. However, due to its invasive nature and the fact that patient management is hardly adapted, it is rarely performed in the clinical workflow. An alternative to IVUS is intravascular optical coherence tomography (IOCT), which features higher resolution at the cost

of lower depth penetration. The main goal of IVUS is to identify thin-cap fibroatheromas (TCFAs). TCFAs feature a large necrotic core surrounded by a thin fibrotic membrane cap and are often postmortally identified as the cause of ACS. However, the authors of reference [Kub10] state that in a 12 month follow up study, only 25% of the originally identified TCFAs remained high risk plaques while the rest lost their vulnerable characteristics, although the mechanism of healing is not explained yet. Moreover, plaque rupture is often clinically silent [Bur01], making it close to impossible to generate labels regarding plaque rupture. In CCTA there are several plaque features related to the vulnerability of plaques proven to increase the patients relative risk for ACS [Puc14] (see Figure 1.2). From these features the NRS and overall calcium burden have already been used in approaches similar to ours as described in [Zre18] and [Kol17]. The first approach was already discussed in Section 1.2.1. It provides a multi-task neural network solution to both plaque and stenosis detection and grading. In order to assess the plaque nature they labeled segments as non-calcified(0%-10%), mixed (10%-49%) or calcified (50%-100-%) based on the proportion of calcified volume of the plaque and produced promising results [Zre18]. The second mentioned approach uses radiomics in order to detect the napkin ring sign (NRS). As this is a qualitative feature the reader dependency of the labels may be high, however it is caused by the difference between the lipid-rich necrotic core and fibrous plaque tissue attenuation [MH10] and therefore a strong hint towards TCFAs.

1.3 Baseline

For the data extraction and labeling an already existing coronary plaque analysis prototype program developed by Siemens Healthineers was used. This program already provided many suitable features:

- CPR and MPR view creation,
- automated centerline extraction and lumen segmentation based on the methods described in [Zhe13] and [Lug14],
- annotation of segments by defining proximal (closer to the vessel root) and distal (further away from the vessel root) reference markers,
- histogram thresholding for tissue quantification enabling the plaque segmentation [Ras06],
- manual segmentation correction capabilities,
- 2D measurements for all views,
- radiomics calculation using the Pyradiomics library [vG17],
- and persistence functionalities.

The provided data collection was created at the Cardioangiologischem Centrum Bethanien (CCB) in Frankfurt am Main in the scope of the medical doctoral thesis of Anika Reideshöfer supervised by Dr. med Joachim Eckert.

1.4 Contribution

Within the scope of this thesis many objectives needed to be fulfilled. Since the labels provided by the clinical collaborator were on branch level only without any information about which therapeutic decisions were taken or considered based on the imaging data, they could not be directly used for our task. Therefore, the whole dataset was re-labeled in the scope of this work with respect to all high-risk plaque features, but the NRS – which could only be identified in four segments. This was done to the best of our ability and based on the measurements provided by the coronary plaque analysis prototype program. Also the labels for segment-wise revascularisation need were crafted based on the branch labels and the measured stenosis degree. In the next step segments were excluded due to movement artifacts, bifurcations and wrong/inaccurate annotations.

Furthermore, this work implemented two different machine learning systems. In order to enable these methods an automatised data extraction was implemented. This included the creation of MPR image stacks for the deep learning approach and the automatic radiomic feature calculation for the radiomics approach, in a batch wise fashion for our whole data collection. The extracted data was then further processed with a standard machine learning setup including feature normalisation and ten-fold cross-validation. The deep learning approach was evaluated for different feature extraction and sequence analysis models. The different feature extraction models were a 2D CNN with average pooling over the centerline dimension, a 2D CNN with fractional max pooling and average pooling over the centerline dimension and a 3D CNN. The sequence models evaluated were an RNN based on gated recurrent units (GRUs) and the WaveNet approach, which is based on 1D convolutions paired with average pooling. All models were also evaluated with respect to the padding method used — zero padding or resizing. The radiomics approach was evaluated for different mask types — segment only, plaque only, plaque and segment — and four different classifiers, which included GNB, boosted SVM, random forests and XGBoost. Both approaches were integrated in the coronary plaque analysis program for the most important target — the need for revascularisation — in order to prove practical applicability of the different algorithms.

1.5 Overview

After introducing the main motivation of this topic together with general concepts linked with this thesis, we set our work in context in Section 1.2, before stating the baseline we are working with and outlining the main contributions in Section 1.3 and Section 1.4. Next we will go into further detail regarding problems faced with the goals of this thesis before setting the medical background foundation for the rest of this thesis. Then the concepts of both machine learning approaches are discussed thoroughly with all connected constituents. Moving on, the characteristics of our data collection will be presented with an analysis of the correlation between different targets in Section 3.1.2, before explaining the concrete implementation of both machine learning approaches and their integration to the prototype. This is followed by a extensive evaluation of both approaches and a comparison between them and related work in Section 4. Finally, we give an outlook of possible new directions with these approaches before summarising this thesis.

Chapter 2

Theory

After introducing our topic, we want to specify all goals of this thesis in more detail, whilst also mentioning associated problems. In order to create a better understanding of the underlying medical issue the genesis of plaque and its structure will be presented. Subsequently, our machine learning approaches will be explained in detail. In Section 2.3 different groups of radiomic features and how they are calculated are explained. Also, the used feature selection method is described, before giving an explanation of the used classifier algorithms: boosted SVMs, GNB, random forests and XGBoost. Moreover, we will introduce the basic theory of deep learning and some advanced structures used in the scope of this thesis. Finally, we will explain the used evaluation metrics and why we used them.

2.1 Goals of the Thesis

Our main goals are:

1. Predict probability of plaque rupture,
2. identify patients and plaques with need for revascularisation and
3. integrate our machine learning approaches to a coronary plaque analysis program.

Goal 1 already leaves us with some challenges. As previously noted in Section 1.2.2 future events are close to impossible to predict based on image evidence. Plaque rupture itself is often clinically silent [Bur01] and even if potential vulnerable plaques are identified, they might actually lose their high risk characteristics, due to medication or healing [Kub10]. Therefore, we choose not to take the probability of rupture as a target, but the high risk features proven to increase the relative

risk of ACS [Puc14].

The second goal of this thesis is derived from our actual labels: we know for each main branch from every patient, if this branch was revascularised or not. Consequently, we try to predict the clinical decision solely based on the image information, knowing that its probably influenced by other factors like overall patient condition, clinical work flow or patient management decisions. Finally, we use an existing coronary plaque analysis program and integrate the machine learning algorithms, which were externally evaluated in order to verify, that the algorithms implemented are robust and fast enough to be implemented in the clinical work flow.

2.2 Medical Background

Before explaining the methods used for differentiating between different plaque types, we want to give some insights about the medical problem this thesis tries to solve. Therefore, we will briefly explain how atherosclerosis arises and how patient management is adapted consequently.

2.2.1 Anatomy

The coronary arteries are divided into three main branches perfusing the heart muscle: left anterior descending artery (LAD), right coronary artery (RCA) and the circumflex artery (CX), which perfuse the heart muscle. The left main coronary artery gets divided into the LAD and CX close to the aortic root, where the left and the right main coronary artery arise from (Figure 2.1) [Cli18].

2.2.2 Plaque Pathogenesis

In general, plaque segments — defined as antisymmetric focal narrowing of the intima (innermost layer of the vessel) — arise due to blood flow induced stress and following inflammatory reactions or from smaller thrombi. The authors of reference [Fus92] proposed three types of pathophysiologic stages of atherosclerotic plaque. First the endothelial cells, which form the inner “skin” of the vessel, are altered with respect to their functionality. Main cause are disturbances in the pattern of blood flow — especially at bifurcations and bendings, which is enhanced by certain risk potentiating substances as cholesterol or chemical irritants in tobacco smoke as well as infections [Ip90].

These type I injuries lead to an accumulation of immune response cells and lipids. These cells release activating substances in order to proliferate platelet adhesion, which is a main characteristic

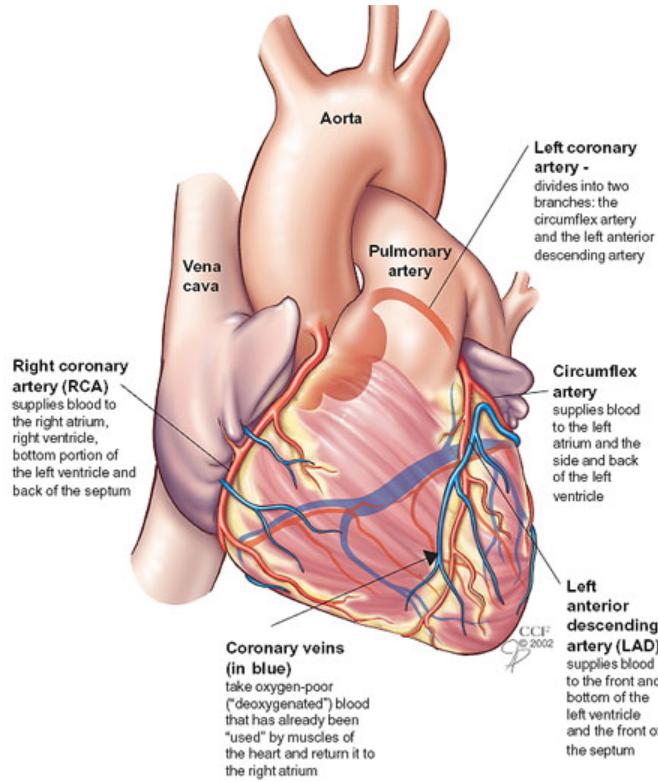


Figure 2.1: Overview over the coronary arteries [Cli18]

of type II atherosclerotic plaque. Both platelets and immune response cells release further growth factors, which cause the segment to grow by additionally attracting smooth-muscle cells, which leads to the formation of a lipid rich necrotic core inside the vessel wall. These lesions are often surrounded by a fibrotic cap and depending on the thickness of this cap, the plaque might rupture due to external stress leading to type III damage.

Plaque rupture leads to thrombus formation, which can have two different results — depending on the size of the thrombus: small thrombi can organize themselves and then lead to a further plaque growth, whilst bigger thrombi can lead to a total vessel occlusion and therefore cause ACS as myocardial-infarction or sudden ischemic death [Ric89] in the case of coronary plaque. These thrombi can also travel into the brain occluding a vessel there, which can cause stroke.

2.2.3 Plaque Treatment

This leads to two different approaches for the treatment of atherosclerosis: medication or angioplasty. So far the treatment with different drugs is mainly focusing on decreasing the amount of

growth factors in order to stop further activation of T-cells and smooth-muscle cell proliferation [Lib02], which can be achieved with anti-inflammatory immunosuppressive drugs. Other than that lysis promoting meds can be used to decrease the risk of ACS based on thrombi.

Revascularisation is most of the time performed by stenting, where in a cath lab a metal grid is guided to the location of the stenosis and then inflated using a balloon. This procedure re-enables blood flow as in a healthy vessel and a healthy perfusion of the heart muscle. However, since stenting creates heavy stress to the endothelia a type I plaque is automatically reestablished and inflammatory reactions to the foreign body lead to a fast atherosclerosis in the same region.

2.3 Radiomics

As medical imaging modalities and algorithms improve, so does the complexity and amount of medical imaging data. The resulting data then needs to be interpreted by a human reader for a qualitative evaluation. However, this discards a lot of quantitative information, which is inherent in the data, but hard to interpret. In order to overcome this problem various quantitative measures of different pathologic characteristics, which are not necessarily perceivable by the human reader, were introduced as radiomics by the authors of reference [Lam12]. First, radiomics were used for the assessment of tumors, since solid cancers are spatially and temporally heterogeneous, making quantitative measures an obvious choice. And as mentioned in Section 1.2.2 plaques of different kinds also have distinct quantitative and qualitative characteristics. Moreover, in [Kol18] radiomics were also used to identify the qualitative high risk plaque feature NRS, proving their applicability for coronary plaque analysis. The workflow to extract radiomics from coronary arteries is sketched in Figure 2.2.

2.3.1 Radiomic Features

In general radiomic features can be divided into four groups:

1. Intensity-based metrics,
2. texture-based metrics,
3. shape-based metrics and
4. transformation-based metrics.

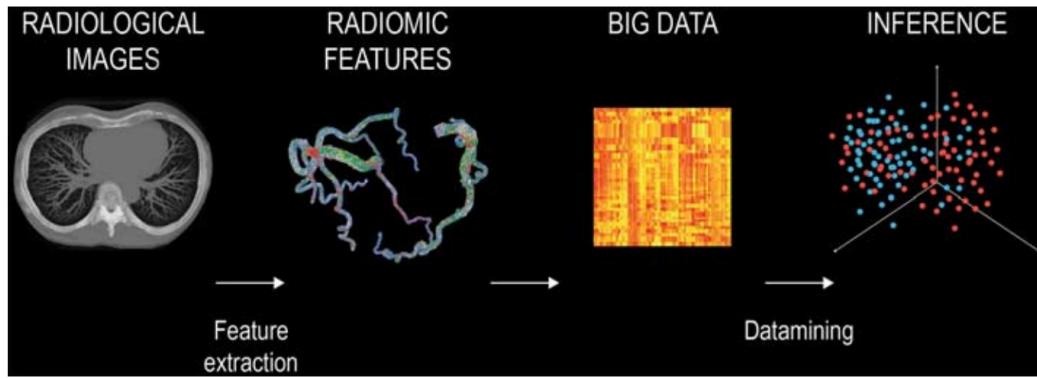


Figure 2.2: Radiomics Pipeline [Kol17]: after segmentation of the the plaque segments a mask is extracted. With this mask different radiomic features can be extracted forming large feature vectors. In order to create feature vectors suitable for the classification data mining with respect to the most discriminating features is performed. These features are then used to build a classifier.

On the following pages we will explain all subclasses of these feature groups with their respective meaning for our goals.

Intensity-based features

The first group of features is solely calculated on the intensity values of the image. Most of these features can be directly derived from a histogram, which counts the frequency of specific image values. Common intensity-based radiomic features are:

- mean: the average gray level intensity,
- median: the median gray level intensity,
- percentiles: amount of values below a certain percentile in the histogram,
- interquartile range: range between the 25th and 75th percentile,
- variance: degree of variation around the mean,
- standard deviation: cut off value in which approximately 68% of the data is located centered around the mean,
- skewness: measure of how asymmetric the values are distributed around the mean
- kurtosis: how close the data points are to the mean,
- energy: overall magnitude of intensities,

- entropy: information measurement and
- uniformity: similarity of all values.

For coronary artery plaque assessment one quantitative metric is already used in the clinic: the Agatston score [Aga90] — defined as the volume of the lesion multiplied by a weighting factor based on the plaques maximum intensity. However, this metric is out of scope for our data at hand since it is calculated using a special native computed tomography (CT) scan. Also, adding quantitative plaque features to the Framingham risk score, has increased the diagnostic accuracy (area under curve (AUC) increase from 0.64 to 0.79) [Ver13].

However, these metrics are often calculated using the absolute HU values, which are proven to be highly dependent on the scanner type used [Wil14]. Therefore, the authors of reference [Kol17] state, that the resulting uncertainty of reproducibility and system and reconstructing algorithm dependency raise major concerns for the use of these metrics. But to overcome these problems, images can be normalised according to some distinct HU scale, or reference values inside of the image (e.g. the mean value of the contrast agent enriched aortic root).

Texture-based features

The next group of features also incorporates the spatial information between neighbouring voxels in order to assess the texture of the lesion. These features are also referred to as second-/higher-order statistics. In order to calculate these features, the gray level co-occurrence matrix (GLCM) needs to be calculated. The steps needed for the calculation of GLCMs can be seen in Figure 2.3. Each entry of the GLCM counts how many intensity values co-occur in a specified direction and distance to the currently considered value. Originally, Reference [Har73] proposed 14 different statistics, which can be calculated using the GLCM. Generally, these are calculated by weighting the entries of the matrix with respect to the property of interest. Examples for radiomic features derived from the GLCM are:

- cluster prominence: measures the skewness of the GLCM,
- cluster tendency: measurement of how many neighbouring voxel values belong to the same group by emphasising the diagonal of the GLCM,
- contrast: measures the intensity variation by emphasising on entries away from the diagonal,
- correlation: linear dependency of the grey level values to their corresponding voxels and

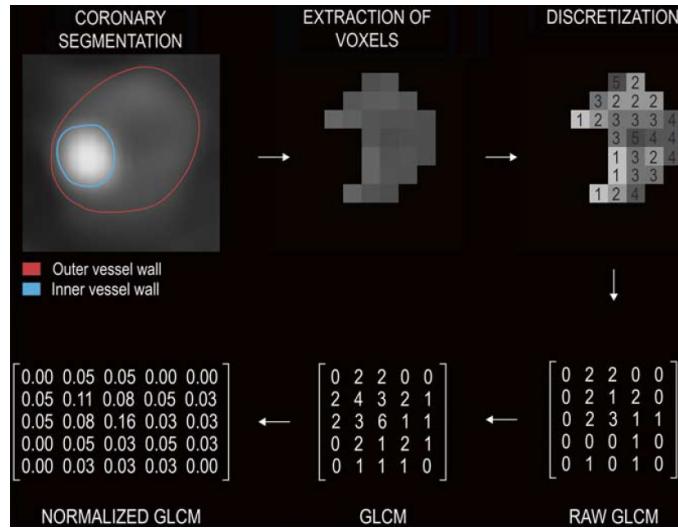


Figure 2.3: GLCM calculation [Kol17]: After segmentation of the inner and outer wall of the lesion, the masked image is extracted. The resulting voxel values are then binned and discretised in order to reduce the amount of different values. For the calculation of the raw GLCM a direction and distance is determined. Then for each appearance of a voxel value j the number of appearances of the voxel value i in the specified direction and distance is counted and inserted to the i th column and j th row of the GLCM. Then the transpose of the raw GLCM is added to itself to create a symmetrical matrix. And finally, the GLCM is normalised with respect to the frequency of the values in order to obtain a normalised GLCM [Kol17].

- homogeneity/inverse difference: inverse of the contrast weights also emphasising the diagonal.

All of these statistics are calculated for the four different GLCMs in order to achieve rotationally invariant statistics. Extending the concept of GLCMs Reference [Gal74] proposes higher-order statistics using the gray level run length matrix (GLRLM). Here, instead of just focusing on neighbouring pixels, the amount of voxels with the same value next to each other are evaluated. The workflow for the calculation of the GLRLM is presented in Figure 2.4. Four angles ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) are considered – analogous to the GLCM calculation – and the amount of specific repetitions j for each grey value i is counted in the single GLRLM before averaging over all angles. Corresponding radiomic features, calculated from the GLRLM are:

- short run emphasis: measures the distribution of short run lengths, corresponds to fine textures,
- long run emphasis: measures the distribution of long run lengths, corresponds to homogeneous areas,

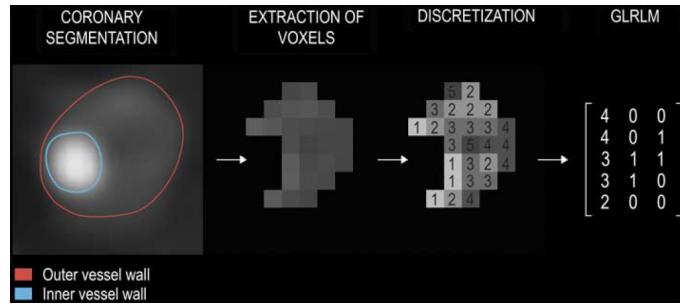


Figure 2.4: GLRLM calculation [Kol17]: After segmentation of the inner and outer wall of the lesion, the masked image is extracted. The resulting voxel values are then binned and discretised in order to reduce the amount of different values. For the calculation of the raw GLRLM a direction is determined. Then for each appearance of a voxel value j the number of consequent appearances the same voxel value in the specified direction is counted and for each amount inserted to the i th column and j th row of the GLRLM where j is the number of repetitions [Kol17].

- gray-level non-uniformity: similarity of gray-level intensity values,
- run length non-uniformity: indicates the homogeneity of run lengths of different intensities and
- run percentage: corresponds to the coarseness of the texture as the ratio between number of runs to amount of voxels.

The underlying concepts of GLRLMs and GLCMs can also be used to create different higher-order statistic features using different rules. Examples for this are the gray level size zone matrices (GLSZM) [Thi13], neighbourhood gray-tone difference matrices (NGTDM) [Ama89] and gray level dependence matrices (GLDM) [Sun83].

Although higher-order statistic radiomics are well established for the assessment of cancer homogeneity [O'C15, Dav12, Par14], they have some major drawbacks when transferred to the task of assessing plaque segments: contrary to tumors coronary artery lesions are very small leading to natural small run lengths and whilst tumors are rotationally invariant, plaque segments are not, since they need to be assessed orthogonal to the vessel center.

Shape-based features

Contrary to above features, shape-based features try to assess the volumetric character of a lesion. Some of these are already well-established in the clinical workflow. These features can also be based on just measuring distances and volumes of the atherosclerotic lesion. Also 2D metrics are well established like for example the area of the severest narrowing in relation to the proposed

healthy lumen. Furthermore, 3D metrics can be used to assess the volumetric properties of the stenosis. These include:

- volume,
- surface area,
- sphericity: measure of roundness of the lesion,
- spherical disproportion: ratio of the surface area of a lesion to the surface area of a sphere with the same volume as the lesion and
- eigenvectors, eigenvalues and associated measures

So far, mainly 1D metrics have been used in the clinical workflow, like lesion length [Cla11] or stenosis severity [Cur16]. But also 2D measurements have been used to describe positive remodeling [Ach04]. Most of these features are easy to calculate and very comprehensible. However, they heavily rely on the segmentation algorithm used and are still mainly focused on their roundness in a spherical sense. Moreover, arteriosclerotic lesions are defined as radial deformations of the vessel along the centerline. Therefore, cylindrical measures would fit this task better.

Transform-based features

Unlike the features discussed thus far, these features were only partly used in the scope of this thesis. The idea of transform-based features is based on the fact that in other domains different features can be calculated. The image data can for example be transformed into the frequency domain without any information loss. Here, the different representation of the underlying data can be used to calculate different features. Possible transformations are for example:

- fourier transform: decomposition of the image information to the frequency space
- wavelet transform: transformation into the frequency domain whilst maintaining spatial information
- log-filtering: emphasising high frequencies and therefore edges.

The use of the frequency domain is widely used in image and signal processing in general and are also already used to analyse characteristics of lesions [Bal14]. However, they need to be treated carefully, since the frequency information of image data might also be vendor specific or

classifiers might overfit due to the large amount of obtained features. Also they are often hard to interpret. Apart from these transformations, various operations can be performed on the masked original image to emphasise/reduce outliers, smooth the image and perform the same calculation of shape-, intensity- and texture-based features on these images, which was also included in this thesis.

2.3.2 Feature Selection

One challenge linked to radiomics is the large size of the resulting feature vectors ranging from hundreds to thousands of individual parameters. Even though a rule of machine learning is “the more data the better”, this only refers to the amount of training samples. A large amount of features bears problems, since irrelevant and redundant features impact the classification process. By selecting specific features before training you can understand which features are really important to get a better understanding of the data, reduce the amount of training time, coping with the curse of dimensionality and improve the classifiers performance [Guy03]. In [Kol17], where a similar approach is described, every single feature was evaluated with respect to its ability to discriminate whether a lesion shows the NRS or not. Individual markers can be identified with this approach, however cross-influences are discarded. Therefore, we chose to extract a subset of features for each classification task, which contains the most discriminative information. Concluding, the best k variables from the feature vectors need to be selected. The metric used for this task was the one-way analysis of variance (ANOVA) f-value, which was calculated for each feature. The F-statistic for one-way ANOVA is defined as:

$$F = \frac{\text{variation between sample means}}{\text{variation within the samples}} \quad (2.1)$$

This means, that we examine if the variation and mean of each feature — when clustered in the separated classes — are significantly different. For each feature mean and variance of features belonging to each separate class is calculated and afterwards compared to the other classes. Three requirements for the use of ANOVA exist:

1. The features are independent,
2. a normal distribution of residuals and
3. the standard deviation over all groups is equal.

The third criteria can be achieved by normalising the data, whilst the second can be assumed for

our data at hand. However feature independence can not be assured. The underlying concept can still be used (as stated in [Low14]), but with a loss of discriminative power.

2.3.3 Classifiers

The selected radiomic features are used to for different classification tasks. A classifier always tries to maximize the probability of a sample \mathbf{x} to belong to the correct class y . This is defined as the a posteriori probability $p(y|\mathbf{x})$. The posterior probability can be expressed by other probabilities using the bayesian rule [Dud12]:

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}) \cdot p(\mathbf{x}|y)}{p(y)} \quad (2.2)$$

Concluding a bayesian classifier selects the class with the highest posterior probability $p(y|\mathbf{x})$. Derived from Equation 2.2 two kinds of modeling exist:

- generative modeling: the class prior probability $p(y)$ and the class conditional $p(\mathbf{x}|y)$ are estimated
- discriminative modeling: $p(y|\mathbf{x})$ is directly estimated

Concluding, a sample with a given feature vector \mathbf{x} is assigned to the class y which it is most probable to belong to. A resulting decision boundary is displayed in Figure 2.5.

Gaussian Naïve Bayes

One of the simplest classifiers is the naïve Bayes classifier [Fri01]. It models the class dependent probability density function based on a very strong assumption: all features are statistically independent from each other. Therefore, they can be factorised as:

$$p(\mathbf{x}|y) = p(x_1, x_2, \dots, x_d|y) \quad (2.3a)$$

$$p(\mathbf{x}|y) = p(x_1|y) \prod_{i=2}^d p(x_i|y, x_1, \dots, x_{i-1}) \quad (2.3b)$$

$$p(\mathbf{x}|y) = \prod_{i=1}^d p(x_i|y) \quad (2.3c)$$

Now, — assuming that all features are Gaussian-distributed — the underlying probability density functions $p(x_i|y)$ can be approximated. Moreover, the prior $p(y)$ can be determined as the

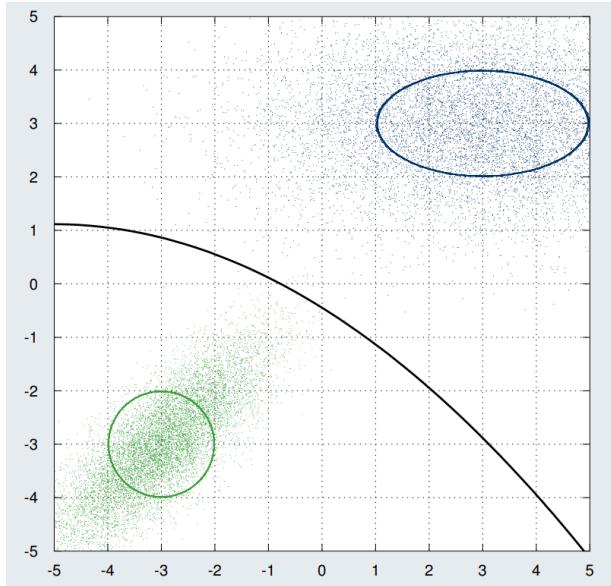


Figure 2.5: Gaussian distributed samples of two different classes – green and blue – in a two dimensional feature space. Between them a quadratic decision boundary [Noe16].

frequency of this class in the given data collection. Therefore, they can be directly substituted to Equation 2.2, in order to create a classifier. The benefit of this approach is, that it performs well on small sets of data and is rather less affected by the curse of dimensionality. However, it does make a strong independency assumption. This drawback can be overcome by efficient feature selection, i.e. by selecting features independent from each other, though.

Support Vector Machines

Another classifier group commonly used are SVMs [Fri01]. SVMs try to find a decision boundary with the maximum distance to both classes. Such a function can be defined as:

$$f(\mathbf{x}) = \boldsymbol{\alpha}^T \mathbf{x} + \alpha_0 \quad (2.4)$$

in the linear case with $\boldsymbol{\alpha}$ as a set of weighting parameters. An example decision boundary is displayed in Figure 2.5. The optimization task for the SVM then comes down to:

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmax}} \frac{1}{\|\boldsymbol{\alpha}\|_2} \quad (2.5a)$$

$$\text{subject to: } y_i \cdot (\boldsymbol{\alpha}^T \mathbf{x} + \alpha_0) \geq 1 \text{ for all } i \quad (2.5b)$$

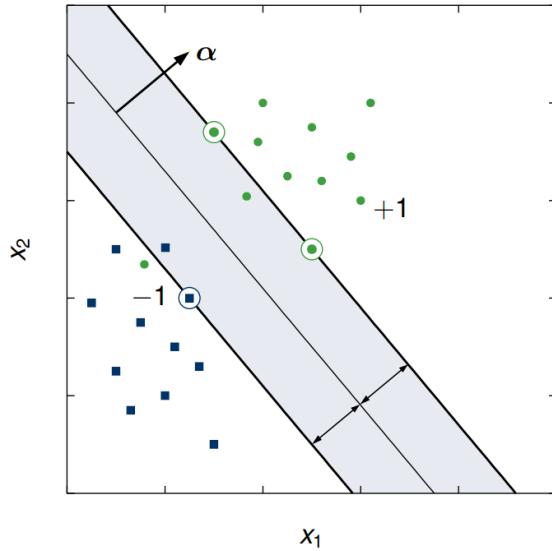


Figure 2.6: SVM decision boundary between two non-linearly separable classes. Samples inside of a margin around the decision boundary are assigned to the right class with the use of slack variables [Noe16].

which can be interpreted as the calculation of the convex hull around each sample set and the maximisation of the distance between them. For non-linearly separable classes and to prevent over fitting, above constraints can be adapted to:

$$\underset{\alpha}{\operatorname{argmax}} \frac{1}{\|\alpha\|_2} + \mu \sum_i \xi_i \quad (2.6a)$$

$$\text{with: } -(y_i \cdot (\alpha^T \mathbf{x} + \alpha_0)) - 1 + \xi_i \leq 0, \quad (2.6b)$$

$$-\xi_i \leq 0 \text{ for all } i \quad (2.6c)$$

where ξ_i denote slack variables, which relax the stiff condition of the original constraint, by allowing some outliers on the other side of the decision boundary. For SVMs further extensions, like polynomial decision boundaries, exist making them applicable for many classification tasks. In the scope of this thesis we used radial basis function (RBF) kernels for our decision boundary. Kernels are functions that map the original feature space to a feature space more fitting for linear decision boundaries. In the case of the RBF kernel this boils down to the dot product being replaced with the kernel function:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (2.7)$$

However, SVMs are exponentially expensive in terms of computation time depending on the amount of features. Since our original amount of features is very high, proper feature selection is also crucial for SVMs.

Boosting

Boosting is often used as a common way to increase the capacity of a classifier algorithm. In general this is performed by taking an ensemble of several weak classifiers, but it can also be used with standard classifiers like SVM. Each classifier is trained with a different subset of the data and under different conditions in order to create a multitude of classifiers. The prediction of this classifier ensemble is a weighted majority vote of all single classifier predictions. The most known popular boosting algorithm is adaptive boosting (AdaBoost) [Fre97]. Before fitting a new classifier for the ensemble, the previously predicted samples are weighted depending on whether they were predicted correctly or not. This enhances the importance of “hard“ cases, since they get a higher weight each time they were misclassified. Then it tries to minimize the loss function:

$$E_t = \sum_i E[F_{t-1}(x_i) + f_t(x_i)] \quad (2.8)$$

where $E(F)$ is any error function, $F_{t-1}(x_i)$ is the prediction of the previous version of the boosted classifier and $f_t(x_i)$ is the newly added classifier.

Random Forests

The random forest approach was proposed in 2001 in [Bre01] and falls under the machine learning sub-field ensemble learning. It constructs a multitude of decision trees and creates a prediction based on the class with the majority vote. To understand this method the underlying concept of decision trees must be clarified first. Decision trees use linear one dimensional boundaries of single features to discriminate between different sample groups on each level. Due to its tree shape the differentiation becomes exponentially fine. A sample decision tree can be seen in Figure 2.7. Even though they have advantages, like, scale and transformation invariance, they often include irrelevant features and are hard to interpret [Fri01]. Random forests perform multiple decision tree fittings, whilst using a different random subset of samples and also a random subset of features for each tree.

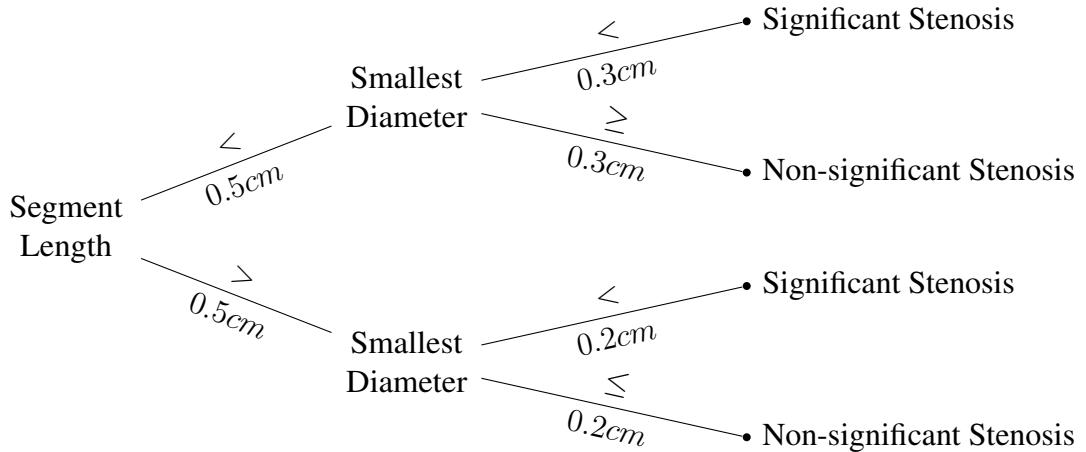


Figure 2.7: Decision Tree

Gradient Tree Boosting

The final classifier we investigated was the gradient tree boosting algorithm, as recently proposed in reference [Che16] under the name XGBoost. It also uses decisions trees, but tries to minimize the following loss function:

$$\mathcal{L} = \sum_{i=1}^n l(y_i, \hat{y}^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t) \quad (2.9a)$$

$$\text{where } \Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (2.9b)$$

In greedy search manner a new tree $f_t(\mathbf{x}_i)$ is added to the ensemble based on how well it improves the prediction. Each tree has an associated number of leafs T and leaf weights w . Moreover, l is a convex differentiable loss function, based on the label y_i and the prediction \hat{y}^t where t denotes the training step. $\Omega(f)$ is a regularisation term, which penalizes the complexity of the model and prevents overfitting by smoothing the learned weights. Since \mathcal{L} is differentiable and convex, it can be optimised with respect to optimal leaf weights w^* . XGBoost gained a lot of attention in recent research, since it outperformed many other classifiers on public datasets on the Kaggle platform [Che16]. It features a high model capacity with sufficient regularisation to prevent overfitting.

2.4 Deep Learning

Deep learning is often linked to the “great artificial intelligence (AI) awakening“ of the past few years. Recently, they outperformed many state of the art classification algorithms. Most notable is the advancement in the ImageNet challenge, where around 14 million images belonging to 1000 classes need to be classified. After the introduction of CNNs in 2012 by Reference [Kri12] the classification capabilities of neural networks exploded resulting in a “super human“ performance by 2015 [He16] — in particular by the introduction of residual networks (ResNets). Apart from their success in image classification, neural network learning approaches have been transformed to many other fields, as object segmentation [Red16], reinforcement learning [Sil16] or sequence analysis [LeC15]. Since the field of neural networks has grown bigger along the years, the term deep learning is often used to describe the process of optimising a neural network. In the field of medical imaging deep learning algorithms also have become well established over the last few years. In the following sections we want to introduce the main constituents and concepts of neural networks, how they are trained.

2.4.1 Constituents

Perceptron

Even though neural networks underwent a rapid development recently, its fundamental idea is quite old. It was proposed in 1958 by Rosenblatt et al. [Ros58] as the perceptron. The Rosenblatt perceptrons decision rule is described by a simple formula:

$$\hat{y} = \text{sign}(\mathbf{w}^T \mathbf{x}) \quad (2.10)$$

stating that its prediction \hat{y} simply depends on the sign of the sum of the inputs \mathbf{x} multiplied by weights \mathbf{w} . However, this approach has one major drawback: due to the linear nature of its decision boundary the XOR function (“exclusive or“) cannot be approximated. An improved version of the original perceptron is the multi-layer perceptron (MLP), which fixes this drawback [Gro82] by introducing hidden layers between the input and the output (Figure 2.8). Mathematically, the output of one layer is the input of the next and a chain-like structure $f^{(3)}(f^{(2)}(f^{(1)}(\mathbf{x})))$ is created, with each layer representing a function f .

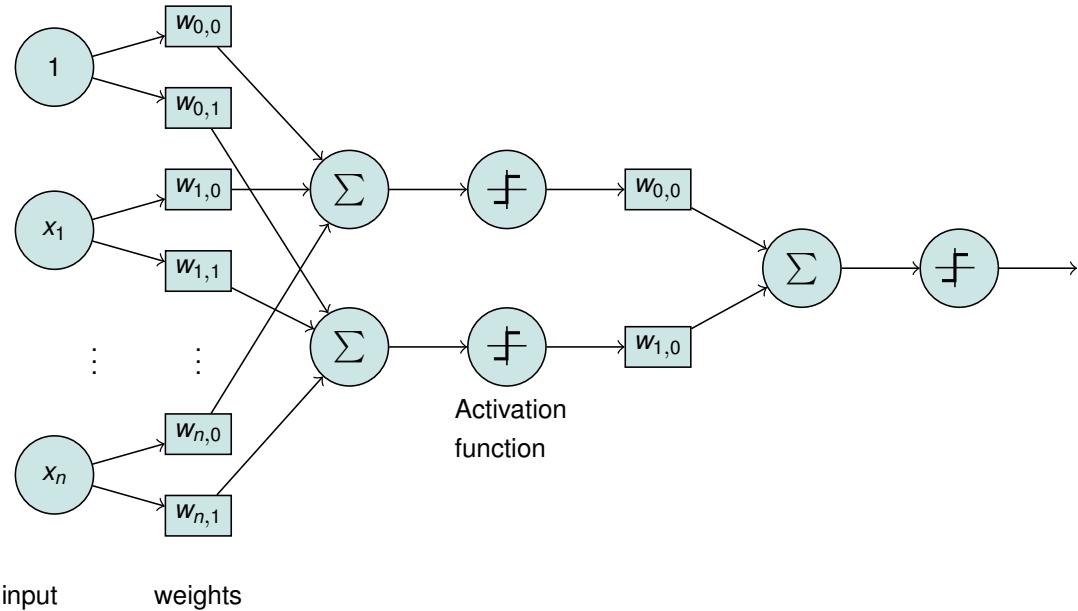


Figure 2.8: In an MLP the inputs are weighted and summed up to create two hidden nodes. The values of these hidden nodes are then again weighted in order to get a final prediction [Mai18]

Learning Procedure

The weights of the MLP need to be optimised such that the network is able to classify correctly. They are randomly initialised, and since all layers formulate a derivable function, the weights can be optimised by calculating the gradient with respect to their role in the prediction. The training of an MLP is performed as follows:

1. The current input of the network is set and propagated forward through the network by multiplying it with the respective weights of each layer and their activation until the final layer is reached.
2. The prediction of the final layer is compared with the labels, all wrong prediction values form an error.
3. This error is then propagated backwards through the network, starting at the final layer. At each layer a gradient with respect to the weights is calculated in order to update the weights. Since the forward pass can be described as a chain of functions $f^{(3)}(f^{(2)}(f^{(1)}(\mathbf{x})))$, the chain rule can be used when propagating backwards through the network.

In general the input of the network consists of multiple samples — a batch — in order to stabilise the gradient descent during training.

Optimiser

The optimiser of the network is defined as the update rule for the weights with respect to their gradient. Stochastic gradient descent is the simplest optimiser and defined as:

$$\mathbf{w}^{k+1} = \mathbf{w}^{(k)} - \eta \nabla L(\mathbf{w}^{(k)}, \mathbf{x}, \mathbf{y}) \quad (2.11)$$

where $\nabla L(\mathbf{w}^{(k)}, \mathbf{x}, \mathbf{y})$ is the gradient with respect to the loss function $L(\mathbf{w}^{(k)}, \mathbf{x}, \mathbf{y})$ and η denotes the learning rate of the optimiser. η needs to be carefully chosen, since too big values will not lead to convergence to the minimum of the loss function and too small values will lead to a very long training time. This problem can be solved, by having a steady learning rate decay or by having an adaptive learning rate.

Moreover, the learning process can be accelerated by using a momentum term. By saving the last gradient of the previous update and combining it with the current gradient, a more stable gradient descent is enabled, which also converges faster.

The most commonly used optimiser is adaptive moment estimation (Adam), which incorporates individual learning rates for all parameters. It weights the gradient for parameters, which are not frequently updated, higher than often updated parameters.

Activations

In Figure 2.8 each layer of perceptrons is followed by activation functions, mapping the output of the neurons to a new range. These activations are needed to create non-linearity, since this enables function approximation, and are also motivated by the biological background of neurons. The original activation function of the Rosenblatt perceptron is the step function, which is problematic, since it cannot be differentiated in a useful way for the back propagation. A differentiable approximation for the step function is either the sigmoid function or the tanh function depending on the range of the mapping (Figure 2.9). While these functions introduce non-linearity, they suffer under the vanishing gradient problem, since they map large regions of the input \mathbf{x} to a small range in \mathbf{y} . By multiplication of multiple small gradients during back propagation, the gradient gets very small when it reaches the first layers. An activation function, which performs better regarding the vanishing gradient is the rectified linear unit (ReLU), which maps all negative values to zero, and is linear for all positive values. This strict mapping of all activations below zero to zero however does not allow any gradient to flow in the direction of the weights, which caused the activation. Therefore, ReLUs can “die“ stopping the propagation of any error. A theoretical fix for this was introduced as the Leaky ReLU [Maa13], which scales down all negative inputs by a

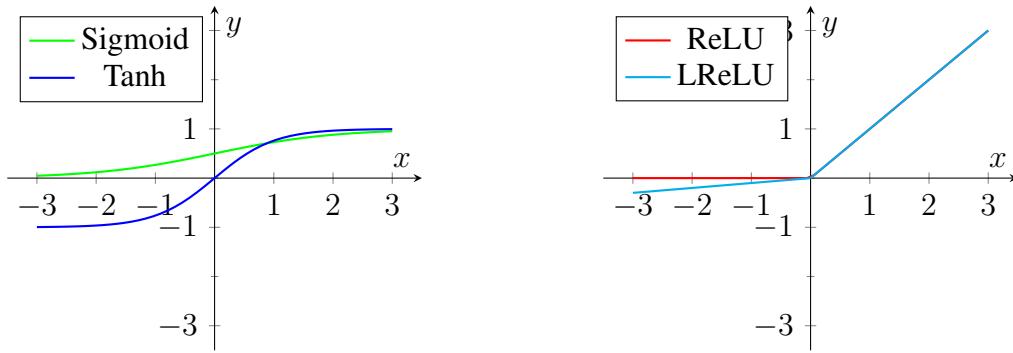


Figure 2.9: Some common activation functions. the sigmoid and the tanh are smooth differentiable step functions within the range of $[0,1]$ and $[-1,1]$ respectively. The ReLU activation is defined as $\max(0, x)$ and also introduces linearity, but sets all negative values to zero. An alternative is the Leaky ReLU, which only scales down negative values [Mai18].

large margin but does not completely discard them in the learning process.

Batch Normalisation

One problem, which originates from the use of ReLUs is the so-called covariate shift. The distribution of the activations of all layers changes according to its parameters due to the ReLU being non-zero-centred. This internal covariate shift can be decreased by incorporating batch normalisation layers, which enforce zero mean and a standard deviation for the activations associated with every batch of data [Iof15]. The use of batch normalisation layers does not only allow a more stable training process but also a faster convergence.

Convolutional Layers

Even though, MLPs are quite useful for classification tasks with feature vectors, approximating any arbitrary function, it does not make sense to use them for images. Any pixel would represent a node and the network would only learn the hard-coded pixel values for each input. Moreover the amount of parameters would explode since the input is two dimensional. In order to overcome this problem, convolutional layers were introduced: the original image is convolved with trainable kernels in order to gain feature maps, which preserve spatial correlations. After obtaining a number of feature maps, their size has to be reduced in order to compress the information and therefore reduce the complexity of the features. This dimensionality reduction is performed with pooling layers, which take the maximum or average value inside of a sliding window with a specified step size (stride), whilst rejecting the rest of the values. A standard network using convolutional layers can be seen in Figure 2.10. A few special cases of convolutions are interesting

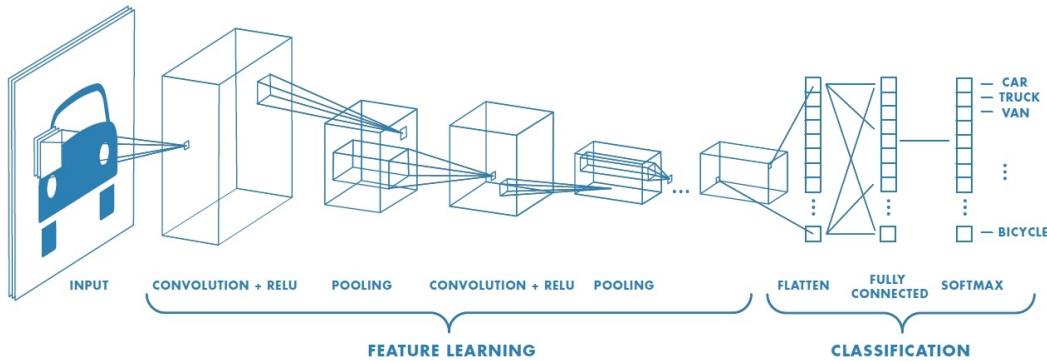


Figure 2.10: Workflow of a CNN. The input image is convolved with several kernels in order to extract features, these features are then compressed by using pooling layers in order to gain more abstract features. These two steps are repeated until the desired feature size is reached. The extracted features are then flattened and a set of fully connected layers weights the extracted features in order to perform classification.

for us:

1. Dilated convolutions: Instead of considering the direct neighbourhood of each pixel during convolution, a dilation rate is introduced defining the distance of the pixels considered for the convolution. This can be seen as a broader field of view, whilst keeping the same amount of parameters.
2. 1D convolutions: These can be interpreted as fully connected layers with coupled weights, therefore the ordering of their input is of high importance. They can be used in order to analyse sequences.
3. 1x1 convolutions: Convolutions with a kernel size of 1 can be interpreted as a linear combination of the previous feature maps.

Recurrent Layers

Another constituent used in the scope of this thesis are recurrent cells. They are most commonly used in order to analyse sequences of data, where the temporal information is of importance. Each time step of the sequence is fed into the recurrent cell at once, while all previous inputs influence an internal hidden state of the cell. The easiest RNN cell is the Elman cell [Elm90], which can be seen in Figure 2.11. Due to the hidden state being mainly influenced by the previous input it does not work well with longer sequences and suffers heavily under vanishing gradients. These problems were partly solved by the introduction of long short-term memory units (LSTMs) by

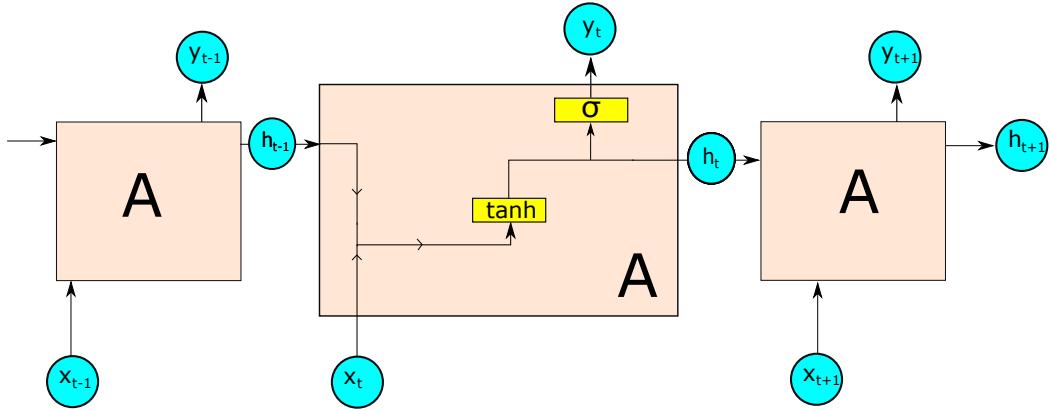


Figure 2.11: Overview over the Elman recurrent cell: each input x_t is combined with the hidden state h_{t-1} from the previous state and weighted to create the next hidden state h_t and the output y_t . Weights are the same for each input of the sequence.

reference [Hoc97]. They use gates in order to access the memory of the cell in form of a cell state additional to the hidden state. The following gates were used:

- Forget gate: weights how much of the old information should be forgotten
- Input gate: decides which inputs are relevant to update the cell state
- Output gate: combines the input with the cell state in order to create the new hidden state and prediction

However, LSTMs suffer under a big number of parameters and are known to be hard to train. Therefore, GRUs were introduced in Reference [Cho14], which keep the idea of having gates control the information preserved inside of the recurrent cell, but foregoes the cell state by just gating the hidden state. Therefore, they have a lower number of parameters while allowing the same performance as LSTM cells.

2.4.2 Advanced Concepts

After introducing the main constituents of deep learning used in this thesis, we also want to give some insight to some uncommon concepts used.

Fractional Max Pooling

The first rather unusual concept is fractional max pooling as introduced in Reference [Gra14]. The most used kind of pooling for convolutional networks is max pooling with a kernel of 2x2

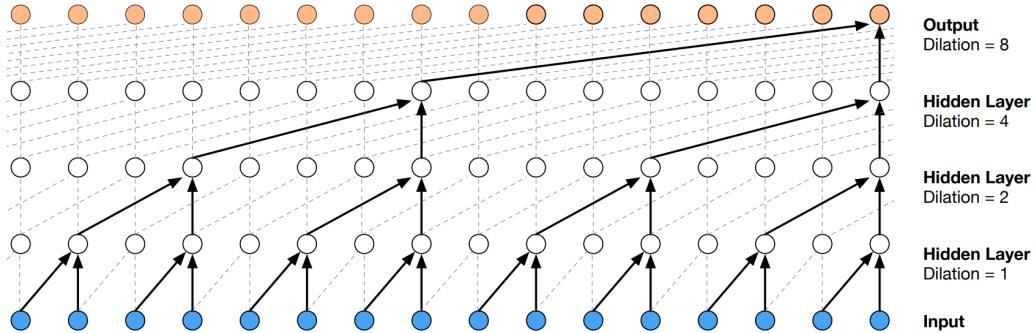


Figure 2.12: WaveNet [vdO16]:The output of WaveNet is based on convolutions with differing dilation rates added together creating a broader FOV during sequence analysis. The resulting information is then further analysed by 1D convolutions and average pooling.

and a stride of 2. This is convenient, since it is not very complex and fast. However, since it reduces the spatial dimensionality so rapidly, the number of feature maps for each abstraction level needs to be very high. Therefore, a decrease rate of $\sqrt{2}$ can be used, in order to create a smoother reduction of dimensionality. The use of fractional dimension reduction is concluding especially interesting, if the starting shape of the image is already quite low. In order to create a pooling operation, which reduces the spatial dimension of the feature maps by a factor of $\sqrt{2}$ the stride is randomly varying between 1 and 2. This allows a deeper network for small inputs or data with a need for a higher amount of spatial abstraction levels.

WaveNet

Recent publications hint that recurrent layers can be replaced by convolutions paired with average pooling [Bai18]. One approach utilising this was published in Reference [vdO16] by DeepMind. In order to gain temporal information between different time steps, 1D convolutions with differing dilation rate are used and combined for each output using skip connections. This leads to the output being a direct combination of all previous time steps (Figure 2.12). After combining different across time dependencies, these are further decreased by using 1D convolutions and average pooling until the target size is reached.

2.5 Evaluation Metrics

To create comparable results, we used a vast number of metrics, which we want to briefly describe, with respect to their meaning for the task at hand. All metrics can be derived by the confusion

matrix defined as:

		Prediction		
		p	n	total
Ground Truth	p'	True Positive	False Negative	P'
	n'	False Positive	True Negative	N'
total		P	N	

which compares the predicted results to the ground truth values. The metrics used in this thesis were:

- Sensitivity, is calculated as:

$$\text{Sensitivity} = \frac{\text{TP}}{P} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2.12)$$

which corresponds to how many of the sick people can be found from the whole amount of sick people in the medical environment.

- Specificity, defined as:

$$\text{Specificity} = \frac{\text{TN}}{N} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2.13)$$

denotes how well healthy people are correctly classified as healthy in relation to all patients without disease.

- The area under the receiver operating characteristic curve AUC is calculated by using different cut-off values, for whether the prediction is positive or negative. This leads to several “operating points” for the sensitivity and the false positive rate (equal to 1 - specificity). The area under this curve is a measure of how good the classifier can be adapted to shift towards a high true positive rate while also having a low amount of false positives occur at the same time.

- Accuracy can be calculated from the confusion matrix as:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{P + N} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (2.14)$$

and is the ratio of the correct predictions to the amount of samples.

- The negative predictive value (NPV), defined as:

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (2.15)$$

is the ratio of true negatives to the number of negative calls. In the clinical workflow a high NPV is associated to the number of really healthy people to the number of suspected healthy people and states how well healthy people can be discriminated.

- The positive predictive value (PPV), is calculated by:

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2.16)$$

is the opposite of the NPV and therefore the more important metric in the clinical workflow.

- The F1 score, defined as:

$$F_1 = 2 \cdot \frac{\text{PPV} \cdot \text{Sensitivity}}{\text{PPV} + \text{Sensitivity}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (2.17)$$

calculates a harmonic average of the PPV and sensitivity. States how well the classifier works for both classes.

- The Matthews correlation coefficient (MCC) can be directly calculated from the confusion matrix as:

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \quad (2.18)$$

Which is used since we mostly deal with very unbalanced class distributions, as it is a quality measure for binary classification, independent from the class balance [Bou17]. Contrary to all other metrics, which are bounded between 0 and 1, the MCC is bounded between -1 and 1.

Chapter 3

Material and Methods

After explaining the theoretical background, we now want to discuss the data and targets used within the scope of this thesis together with related problems. Furthermore, our labeling approach is outlined. Next, we present how we implemented the radiomics and deep learning approaches used for the later evaluation. The necessary preprocessing and data handling steps for both types of approaches, are also described. Finally, we will briefly show how we integrated both machine learning systems into a coronary plaque analysis prototype software solution.

3.1 Data

Data used in this thesis was acquired by CCB in Frankfurt am Main in the scope of the medical doctoral thesis of Anika Reideshöfer supervised by Dr. med Joachim Eckert. This data collection contains CCTA scans of 105 patients taken within a time span of 2 years. All scans were taken with the same system with a slice thickness of 0.5mm and a pixel spacing of 0.3mm. These patients had a total of 463 lesions and were redrawn on site by an expert with a new version of the coronary plaque analysis software solution containing functionality as described in Section 1.3. For each lesion a proximal and a distal marker were set. According to these markers a segmentation for the inner and outer wall was suggested, which was adapted if needed using the histogram and contour correction functionalities. The original meta data contained information about the patient risk groups and which of the three coronary main branches was revascularised. However, no labels on plaque level were given. To address this the data collection was relabeled for the work undertaken in this thesis, considering the following new targets:

- Stenosis degree over 50%: the area of the narrowest point of the lesions was compared to the area for a suggested healthy vessel. This approximation was performed by taking the

area at the proximal and distal marker and weighting their average based on the distance to the point of measurement.

- Positive remodeling: a similar approximation as for the stenosis degree was used. However, the area of the cross-section of the outer wall was of interest now. A remodeling index was calculated as the relative change from the suggested area to the actual area. A cut-off value of 1.1 (area is 10% bigger than the suggested) was chosen, which is the suggested value in reference [MH14].
- Low HU attenuation: This value was directly read from the plaque's histogram. A cut-off value of 30HU was used according to the literature [Mot07]. However, it is not specified which scanner type was used in literature, which can impact the HU scale as described previously in Section 2.3.1.
- Spotty calcifications: Small round calcified volumes within the plaque with a maximal diameter of 3mm were identified as spotty calcification according to the literature [Fer12, Mot09, Mot07]. Note that one segment can have multiple spotty calcifications.
- Revascularisation: The branch level labels were combined with the stenosis degree in order to create labels for the need of revascularisation on the plaque segment level.
- Edgecases: In order to evaluate “hard cases“ separately segments with a stenosis degree between 40% and 60% were marked as edgecases.

Even though most of the targets are based on manual measurements done on site, an additional expert reader would improve the validity of these labels and shall be considered in further studies.

3.1.1 Data Analysis

In order to get a better understanding of how our data was shaped and in order to gain some prior clustering we analysed our data prior to evaluating them with our machine learning algorithms. For each target a different ratio of true positive to true negative values needs to be considered, impacting the classification tasks at hand.

In general the patients had an average of 4.7 plaque sections with a standard deviation of 3.1, while the maximum number of lesions in a patient was 13. A detailed overview over the co-occurrence of all targets is presented in Appendix A. Following observations can be extracted from these Tables:

- The stenosis degree and the need for revascularisation are highly correlated. However, while only 8 segments with stenosis degree greater than 50% were revascularised, 21 segments

were revascularised even though they did not fulfill the cutoff value of greater than 50%. This might be due to ill-defined stenosis degrees as stated in Section 3.1.2.

- Only half of the significant stenosis and revascularised segments exhibit positive remodeling.
- Spotty calcifications are more often associated with revascularisation than with stenosis degree.
- Most segments which were not revascularised but had a severe stenosis grade had negative remodeling.
- As would be suspected, plaques with low HU values and spotty calcifications do rarely occur together.
- Given the low frequency of low HU attenuation plaques they are very often represented in both revascularised and high stenosis grade plaques.
- The most frequent target is positive remodeling followed by spotty calcifications.
- There is a class imbalance for both revascularisation and stenosis grade.
- There is a severe class imbalance for low HU plaques. This might be due to our strict cut-off value.

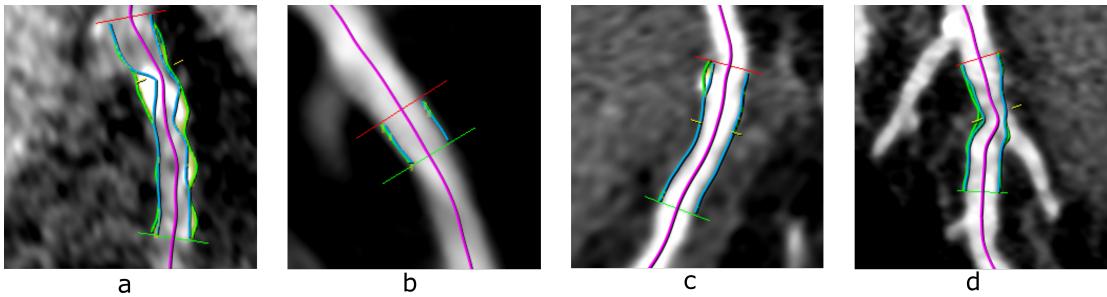


Figure 3.1: Some of the segments needed to be excluded. Movement artifacts and noise prevent an exact segmentation of the plaque (a). Small lesions with no apparent plaque volume were also excluded (b) as well as wrong annotations, where the reference markers do not mark the starting and end point of the plaque segment (c). Furthermore, bifurcations cause the stenosis degree and the remodeling index to be ill-defined (d)

3.1.2 Data Cleansing

For various reasons we could not include all of the data given. Some samples contained heavy noise or wrong/inaccurate segmentations, whilst others had no relevant segments at all. The severest problem with respect to our targets was that the stenosis degree and positive remodeling were often ill-defined. These targets rely on the area at the proximal and distal reference markers. However, if there is a bifurcation between these markers, a natural decrease of diameter occurs. Also for some segments with heavy plaque burden the reference markers were set at different plaques, which also makes them insufficient to be a reference. Some examples of excluded plaques can be seen in Figure 3.1. We excluded 16 segments due to irrelevance, 47 segments due to inadequate annotations and 46 segments due to bifurcations. Therefore, 363 plaques for the task of classifying the stenosis degree and the remodeling and 400 lesions for all other targets were used.

3.2 Radiomics Based Approach

3.2.1 Preprocessing

Segmentation

In order to extract radiomics features we need the segmentation of the outer and inner vessel wall. Luckily, a segmentation algorithm for both was already integrated into the coronary plaque analysis software solution and therefore each sample had the segmented walls saved as persistence data. However, we tried different masks for the feature extraction: Only the plaque, only the

lumen and the segmentation of the lumen and the plaque combined. These were also already obtainable from the prototype without much effort.

Aortic Mean Normalisation

As mentioned in Section 2.3.1 intensity-based features get heavily influenced by the type of scanner used. Moreover, also the tube-voltage has an influence on the HU value distribution. To overcome this problem, the data needs to be normalised to some fixed value. We chose to normalise by the mean of the aortic root. For this a sphere with fixed radius of 3mm was segmented in the middle of the aortic root. Then the mean HU value inside this sphere was calculated and used to normalise all voxels according to:

$$I^* = I - \mu_{aorta} + 1400 \quad (3.1)$$

setting every aortic root mean value to 1400. The mean of the aortic root depends on the attenuation of the contrast agent used during the scan and should therefore be equal across different scans.

Data Extraction

In order to automate the data extraction a batch script was used. This script opened the current dataset and then had to select every single plaque and calculate the radiomic features. The PyRadiomics library [vG17] was used for the feature calculation, which already included all necessary radiomic feature arithmetics. All possible features provided by this library were calculated under Wavelet transform and with the input being squared, square rooted, logarithmised and exponentialised. For each data set one file was created containing the feature vectors of all plaques associated to the patient. Next all data points were read in again and mapped to their respective labels using the patient id and the plaque id. For the branch and patient wise classification task the mean and variance of all feature vectors contained in the respective patient/branch were calculated and concatenated in order to form the feature vector for the branch-/patient-wise classification.

3.2.2 Radiomics Classification

The classification was performed using ten-fold cross-validation. For this the data was split into 10 subsets with 90% training data and 10% test data. The train-test split was performed using stratified group k-fold splitting. For each subset two criteria were enforced: since segments within one patient are per definition correlated with each other they always belong to the same split and

the class distribution is roughly the same between training and test sets.

For each training set the mean and variance for every feature was calculated. These were used to normalise the features of both training and test set using the formula:

$$\tilde{\mathbf{x}} = \frac{\mathbf{x} - \mu_{x_train}}{\sigma_{x_train}} \quad (3.2)$$

Then the feature selection method described in Section 2.3.2 was used in order to reduce the number of features for the classification.

Four classifiers were trained with these feature vectors (see Section 2.3.3): boosted SVM, GNB, random forest and XGBoost. The SVMs was used with RBF kernels and combined with the AdaBoost algorithm. Then the test set was used in order to predict the respective class and the predictions were compared to the associated labels by calculating following metrics: AUC, accuracy, F1 score, PPV, NPV, sensitivity, specificity and the MCC. All predictions and labels were also saved for a final evaluation using all samples. Additionally, the edge-cases were evaluated separately for the revascularisation target. The classifier and feature selection implementation of the scikit-learn python package (version 0.20.0) was used.

3.3 Neural Network Based Approach

3.3.1 Preprocessing

As for the radiomics approach, the HU scale needed to be adapted for different tube voltages. Therefore, also the mean value of the aortic root was used to normalise the image data.

However, contrary to the radiomics approach, we do not need any segmentation here. Instead we use stacked MPR images of the segment: for each centerline point between the two reference markers a plane orthogonal to the centerline is interpolated. The size of these images was set to 60x60, with a FOV of 10mm. This lead to an MPR pixel spacing of 0.16mm which is significantly higher than the pixel spacing of the data (0.5mm slice thickness, 0.3mm spatial pixel spacing). Therefore, the image data was upsampled using trilinear interpolation. This window was chosen to allow rotations and translations in the later pipeline while being able to display the maximum diameter of vessels including positive remodeling.

Padding

Even though it is possible to train neural networks with variable input sizes, this requires the training to be performed on single samples instead of batches. This is highly disadvantageous, since it increases both training time and convergence stability. Therefore, usually all inputs are padded in order to have the same size.

In this thesis we use and compare two different kinds of padding. The first approach is the default method for padding – adding images with only zeros to the stack until the target size is reached. Alternatively we resize the plaque segment by stretching it in centerline direction. The centerline dimension was 170px based on the largest plaque segment in the data set plus a tolerance margin of 10% corresponding to a length of 34mm.

Slicing

Due to the volumetric nature of the resulting plaque image stack a direct 3D convolutional network for feature extraction would be possible. However, since the main characteristics defining the targets should be radial to the centerline and changing across the centerline dimension, the original image stack was sliced into cubes. These had a size of 36x36x16 with an overlap of 8 resulting in 40 cubes. As noted before the MPR image stacks had a spatial dimension of 60x60, so a reduction in x and y direction is performed as well. Note that different cube sizes were evaluated but these

performed best for most targets.

Data Augmentation

Neural networks are known to be very data hungry. Since we only have a very limited data set we artificially augmented our data set. All training images had a chance to be rotated between 0° and 360° in steps of 10° , flipped in the x or y dimension, or shifted by an offset of 3px around the center in order to both extend the variance in the data set as well as creating robustness to inaccuracies. Furthermore, minor gaussian noise with a mean of zero and a standard deviation 0.01 was applied to the images.

3.3.2 Deep Learning Classification

Model Overview

The above described image cubes were processed as in the pipeline shown in Figure 3.2. First features from the single cubes get extracted. For this three different CNN architectures were used: a 2D CNN with conventional max-pooling and average pooling across the depth dimension, a 2D CNN with fractional max-pooling and average pooling across the depth dimension and a 3DCNN. The extracted features were then analysed in a sequence-like fashion using GRUs and a WaveNet architecture. All models had roughly the same amount of parameters (~ 50000). All activation functions apart from the final layer were Leaky ReLUs, due to the known good performance of ReLUs and the theoretical benefits of not punishing below-zero activations too hard. Consequently, batch normalisation layers were used in order to decrease internal covariate shift. The probability mapping in the last layer was performed with the soft-max activation. The optimiser used was Adam with default hyper-parameters, since we have a multitude of different layer types, and binary crossentropy loss for all targets. The model setups can be seen in Appendix B. The code for all models was written with the Keras package(version 2.15). The recurrent convolutional neural network (RCNN) implementation was heavily inspired from the approach described in [Zre18] and the WaveNet implementation is based on the code in this kaggle kernel: <https://www.kaggle.com/kmader/quickdraw-with-wavenet-classifier>. However both implementations were heavily altered.

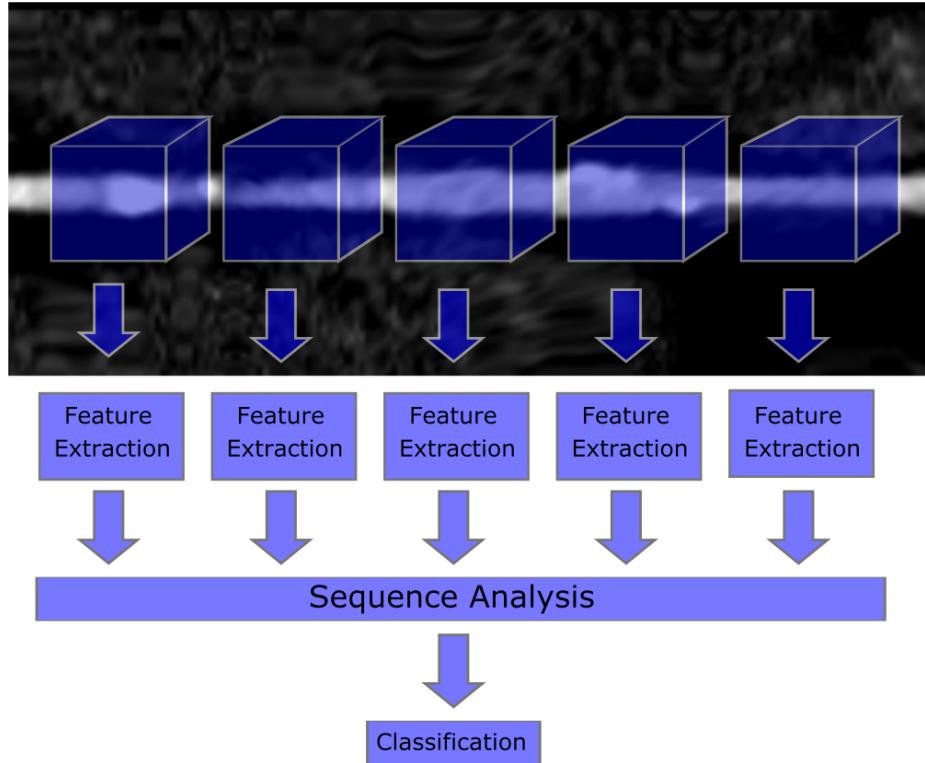


Figure 3.2: Workflow of the Deep Learning Classification

Training Procedure

Analogous to the radiomics approach, ten-fold cross-validation by splitting the data into stratified subsets, while keeping data from the same patient in the same set. Cross-validation is not commonly used for deep learning, since it is not guaranteed that the training converges to the global optimal solution based on the initialisation. However, this work rather wants to present a proof of concept and to validate our results for such a small data set, we needed the approach to perform well on different splits of data. In order to establish class balance during training, the data was re-sampled during batch creation, having each batch contain the same amount of true positives and true negatives. An epoch was defined by having trained the model with each sample at least once. Each model was trained for 50 epochs for each fold, where the weights with the highest MCC during validation were saved and reloaded after training. MCC was used as the metric to optimise since it is known to be a good metric to judge how well a classifier performs even when the class distribution is imbalanced [Bou17]. Since we force class balance by resampling our data during batch creation we enforce both classes — identifying that a plaque has the high risk plaque feature or not — to have the same importance for our model. The intermediate predictions and

re-sampled labels were stored in order to perform a final evaluation. As before, edge-cases were again separately evaluated for the revascularisation target. Training was performed on a NVIDIA DGX-1, which has 8 NVIDIA Tesla V-100 graphics cards, dual 20-core Intel Xeon E5-2698 v4 CPU with 2.2 GHz and 512GB of DDR4 RDIMM RAM and took an average of 10 seconds per epoch for all models, but the ones including fractional max pooling which took 36 seconds per epoch.

3.4 Prototype Integration

In order to prove applicability in the clinical work flow both approaches were integrated into the coronary plaque analysis suite for the specific target of predicting revascularisation using the MeVisLab framework.

Radiomics Approach

First three classifiers needed to be trained for each abstraction level (patient, branch, segment). For this 80% of the data was used in a random fashion. In order to normalise new data, the mean and standard deviation of the training data was saved together with the classifier object.

The classification process is started by a single button press. The feature processing pipeline remained as described in Section 3.2.2. A MeVisLab module using an external python kernel was used in order to ensure the same package version, containing a script with the same processing steps as before but using the saved mean and standard deviation. For each target a prediction was saved and then routed to a graphical user interface (GUI), presenting the predictions in a hierarchical tree-like fashion.

Deep Learning Approach

Due to the small amount of data only one abstraction level can be applied for the deep learning approach. However, the functionality to integrate all three was integrated into the prototype. As before, the data extraction methods of the batch executor were reused and an external python kernel loaded the trained model with the corresponding weights. The data pre-processing was performed as described before and the predictions were displayed in the same format as for the radiomics approach.

Chapter 4

Results

4.1 Evaluation

The methods outlined previously were evaluated and their performance is presented in this chapter. For both approaches different side factors were evaluated – the mask type for the radiomics and the kind of padding for the deep learning approach. Also different classifiers and models are compared for all different targets. Then, our results are put into perspective by looking at state of the art results for other approaches.

4.1.1 Radiomics Based Approach

First, we want to present the results for the radiomics approach. Note that we will focus on the best results in order to increase the information density of this chapter, all tables can be seen in Appendix C.

Best Mask for each Target

The classifiers were evaluated with different masks for each target: lumen only, lumen together with plaque as well as plaque segment only. For different targets different masks worked best:

- Stenosis degree over 50%: Here the lumen only mask performed best, which is not surprising since all of the information of the degree of narrowing should be contained in the lumen itself.
- Positive remodeling: Since this target mainly shows its characteristics in the plaque itself the plaque only mask yielded the best results.

- Low HU attenuation: Here, the difference intensity values of the plaque itself were the most important.
- Spotty calcifications: This is the most surprising result, since the lumen only mask worked best, even though the structure of relevance is not actually in the masked image. Possible explanations for this might be that relatively short segments with high calcification degree were labeled as spotty calcification and were very frequent in the data set. Other than that the high calcification might result in shadowing effects in the lumen and result in inhomogeneities in the lumen. However, all other masks performed very well, too.
- Revascularisation plaque level: Since revascularised plaque segments were identified by selecting those segments, which were in stented branches and exhibit a high stenosis grade, as with the stenosis degree target, the mask of just the lumen performed best.
- Revascularisation branch level: Here all masks seemed to work equally, with the combined mask providing marginal improvements in performance.
- Revascularisation patient level: Contrary to the plaque level revascularisation, the plaque only mask performed best. This is an indicator that other features were important for the risk of the whole patient. Also note, that some masks only differed in the AUC but all other metrics stayed the same, indicating that the same feautures were important and the same confusion matrix resulted from the classification.

Best Features for each Target

For each target and each mask different features were selected by the feature selection algorithm. A list of the selected features for each target is provided in Appendix C.2. In summary these feature groups were selected for the masks stated above for their respective target:

- Stenosis degree over 50%: Since the lumen only mask was performing best, mainly features corresponding to the shape (compactness, elongation, flatness, diameter, sphericity, area,...) of the lumen were selected. This is logical since there are very little intensity changes within the lumen itself.
- Positive remodeling: Even though this target should also rely on the shape of the segment, the selected features mainly involved intensity based features, with exponential filtered second order statistics being the most prominent.

- Low HU attenuation: Several different second order statistics such as long run and short run emphasis performed best.
- Spotty calcifications: Here surprisingly the lumen only mask was the best choice and also most metrics are shape based with the exception of the gray-level non-uniformity metric calculated from the GLSZM. This metric might be corresponding to shadowing effects.
- Revascularisation plaque level: Analogous to the stenosis degree, the shape based metrics were mainly chosen by the feature selection algorithm, with the exception of the correlation feature of the GLCM.
- Revascularisation branch level: For the branch level a mix between shape-based and second-/first-order statistics performed best.
- Revascularisation patient level: In order to predict the need for revascularisation on the patient-level, second-order statistics paired with the wavelet transform performed best.

Classifier Comparison

To classify the samples and their corresponding feature vectors four different classifiers were used: boosted SVM, GNB, random forest and XGBoost. These where evaluated for all targets with above specified features and masks. In general, boosted SVMs, XGBoost and random forest classifier (RFC) all should have the same capabilities: they can solve the XOR decision function, they should generate a non overfitted decision boundary (boosted SVMs are the most prone to overfitting) and generally create a very good classifier with the given features. GNB on the other hand is not able to approximate the XOR function, and is based on a very strong independency assumption as well as the assumption that the samples of each class are gaussian distributed. A good performance of the GNB classifier therefore suggests that the underlying problem is rather simple and these assumptions hold true. With respect to the different targets following classifiers performed best:

- Stenosis degree over 50%: The RFC performed best in most metrics – but not significantly better than the other classifiers – with an AUC of 0.91 and an MCC of 0.59.
- Positive remodeling: For this target all of the classifiers performed rather bad with an AUC of around 0.6. Reasons for this might be the cut-off value used for labeling or the labeling procedure itself, because very often the remodeling index calculated from the prototype did not match the visual perception of the segment. This might be based on ill-defined reference

markers or the calculation procedure used internally. The best classifier for most metrics was the GNB with an AUC of 0.58 and an MCC of 0.32.

- Low HU attenuation: Here the GNB classifier performed best for many metrics, although not significantly. The results were also very good with an AUC of 0.96 and an MCC of 0.65. This is not surprising since the underlying problem for this target is straight forward.
- Spotty calcifications: As mentioned before, surprisingly the lumen only mask performed best for this target, hinting that very small segments, which might not be identified as plaques with spotty calcifications by an expert reader – because the whole plaque consists of one spotty calcification– , were relatively prominent in our data . The results still look quite good for this target, with an AUC of 0.86 and an MCC of 0.69.
- Revascularisation plaque level: Here, most classifiers performed equally. The fact this includes the GNB classifier hints that the underlying problem holds the strong assumptions of feature independence and gaussian sample distribution. In order to set our methods in context we compared them with the stenosis degree as a reference measurement. However, since our labels are partly based on this measure, this reference measure outperformed our approach by a slight margin. However, good scores were reached (AUC = 0.81; MCC = 0.42)
- Revascularisation branch level: On the branch level boosted SVM was the best classifier by a small margin. It scored an AUC of 0.75. and a MCC of 0.39.
- Revascularisation patient level: For the identification of patients with the need of revascularisation the performance was quite similar to the other revascularisation targets(AUC = 0.79; MCC = 0.41) but the sample set was more balanced, leading to different metrics performing better.

4.1.2 Neural Network Based Approach

Analogous to the radiomics based approach, we will not include all the tables in this chapter, but in Appendix C. Here, we want to focus on the performance difference of the padding methods and models used. In general, deep learning algorithms need a lot of data to perform well. Due to this we excluded the branch and patient level revascularisation targets from the final evaluation, even though experiments for these targets were performed as well.

Padding

As described in Section 3.3.1 we used two different sorts of padding: zero-padding and stretching. While better performance of zero-padding hints toward the fact that the preservation of the spatial dimension in the centerline dimension is important, better performance of the stretching hints towards the independence of the target from this dimension. Moreover, stretching instead of padding prevents the network from overfitting on the length of the segments. For different targets, these methods performed better or worse:

- Stenosis degree over 50%: Both padding methods performed equally. Therefore, no conclusion can be drawn.
- Positive remodeling: Here, the stretching in centerline direction performed significantly better. This relates to the centerline dimension being unrelated to the target but more the global change in volume of the plaque.
- Low HU attenuation: Again the stretched inputs performed significantly better, hinting that the global information of having certain HU values is better preserved.
- Spotty calcifications: Since the information about the plaque and calcification length in centerline direction is of high interest for this target, the zero-padding approach performed better.
- Revascularisation plaque level: As for the stenosis degree both padding methods performed equally.

Model Comparison

As described in Section 3.3.2 three different feature extraction and two different sequence analysis methods were used with different capabilities. Concluding, different models performed better on different targets. As a general remark, sequence analysis with the “WaveNet“ was proven to exhibit greater stability during training than training with GRUs. However, the impact of having zero-padded vs resized images was pretty low for the WaveNet sequence analysis approach.

With respect to our targets following networks performed best:

- Stenosis degree over 50%: 3D WaveNet performed best, but not significantly better than the rest. Almost all models scored pretty good metrics with the best having an AUC of 0.89 and an MCC of 0.58.

- Positive remodeling: the conventional 2D and 3D feature extraction paired with GRUs performed best, but still this target was not predicted very robustly with an AUC of 0.69 and an MCC of 0.26. Similar reasons as for the radiomics based approach apply.
- Low HU attenuation: for this target the 3D CNN with GRUs had the best metrics, but not significantly better than the WaveNet approach or the 2D fractional max-pooling (FMP). The good performance (AUC of 0.89; MCC of 0.48) of the deep learning approach for this target is a bit surprising due to the low amount of true positive samples in our data (30 out of 400). Given a better class balance or a higher frequency of these samples an even better score can probably be reached.
- Spotty calcifications: all models performed equally well. As suspected in the radiomics part, a good classification of spotty calcifications might already be reached by predicting the length of the segment, since the frequency of very small calcified segments is very high in our data. This probably led to a convenient local minimum where most of the models got stuck.
- Revascularisation plaque level: contrary to all other targets the best performing model was the 2D FMP. This indicates that the spatial information loss from max pooling does impact the feature extraction capabilities of the other models. The separate evaluation of the edgecases exhibited little correlation with the classification of the rest of the plaques with really bad scores (AUC of 0.60; MCC of 0.14).

4.1.3 Comparison between Deep Learning and Radiomics

Before going into a detailed metric comparison between the two approaches performed in this thesis, we want to recall the major differences between them: the radiomics approach expects segmented inner and outer vessel walls and uses these to compute a large set of features, from which only the best get selected and used for a classification. On the other hand the neural network approach only relies on the centerline extraction and interpolated MPR images along this centerline, and predicts the class of each segment based on a CNN feature extraction followed by a sequence analysis network. Concluding, the radiomics approach does discard a lot of image information by masking in comparison to the neural network approach while on the other hand it uses prior information given from the segmentation. In Figure 4.1 the best models/classifiers for each target are shown. In general the radiomics approach performed slightly –but not significantly – better. A very possible reason for this is ten-fold cross-validation is less applicable for deep

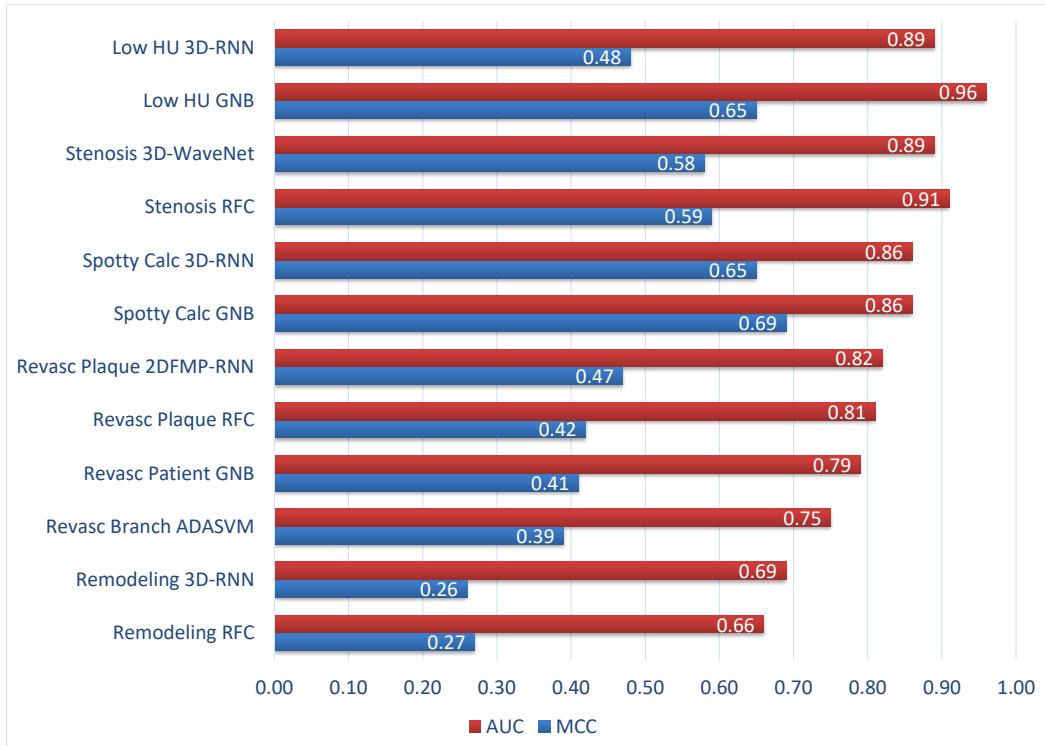


Figure 4.1: Graph of the results of the best performing methods within both approaches

learning, since neural networks do rarely approximate truly convex functions and therefore can get stuck in local minima. Also the data-set used for this thesis was really small for deep learning. Given these limitations it is rather surprising how well this approach performed.

The only target where the deep learning approach performed significantly better was positive remodeling. This target was the overall worst performing, based on the labeling method used. A better performance of the neural network approach suggests that the masking operation in fact decreased the image information with respect to the target.

4.2 Comparison to Related Work

Since the assessment of most targets used for this thesis are based on measurements employed in the clinical practice, automated classification of high risk plaque features is not the main focus of

current research. However, the automated assessment still enables a faster workflow in the clinic. Moreover, branch and patient level revascularisation has to our knowledge, not been considered as a target in any other publication so far. One might argue, that branch or patient wise evaluation of these targets exists, but these neglect the global interplay of the features calculated in this thesis. The first target for which related work exists is the stenosis degree. As stated in Section 1.2 reference [Zre18] proposed an RCNN approach similar to ours. The differences between our and their approach are that we tried different feature extraction and sequence analysis models, had a smaller data set, different cube dimensions and performed cross-validation. Moreover, the model they used had more parameters (~ 350000) and was used to perform multi-task learning: the same feature extraction method was used to assess both calcification and stenosis degree but with two different classification parts. Also they used one network to both analyse and detect segments of interest. With their approach they scored an accuracy of 0.80 and an f1-score of 0.75. The high f1 score coupled with the relatively low accuracy suggest that they had a better class balance in their data set compared to ours. However, no similar metrics other than that were used for the classification part. Contrary, we scored an accuracy of 0.87 and an f1-score of 0.66 with our deep learning based approach and an accuracy of 0.85 and an f1-score of 0.68 for our radiomics based approach. As a general performance measure they used Cohen's κ contrary to our MCC which scored 0.68, while we scored 0.58 for our deep learning approach and 0.59 by using radiomics. Note that these metrics cannot be directly compared. In summary, our method performed at least comparably well given the difference in data set.

Since revascularisation decisions today are mostly based on FFR values we compared ourselves with simulated FFR, described in Section 1.2. In this field many publications exists, therefore we use an overview provided by reference [Tes17] citing eight different studies on this subject and compare ourselves to the range of metrics between those [Nør14, Nak13, Koo11, Ren14, Coe14, Yan16, DG16, Kru16]. All of these studies scored an AUC, PPV, NPV, sensitivity and specificity between 0.79-0.93, 0.56-0.74, 0.84-0.95, 0.76-0.89 and 0.61-0.95, while our our deep learning approach scored 0.82, 0.53, 0.91, 0.67 and 0.85 and our radiomics approach scored 0.81, 0.62, 0.87, 0.44 and 0.93 as visualised in Figure 4.2. Here we must note, that we do not compare ourselves to the reference standard FFR value, but to a self crafted revascularisation label. Also all of their scores are vessel based, whilst we try to create labels based on single segments. Direct comparison is therefore not feasible. However, we want to state that we lie in the range of these methods with most of our metrics.

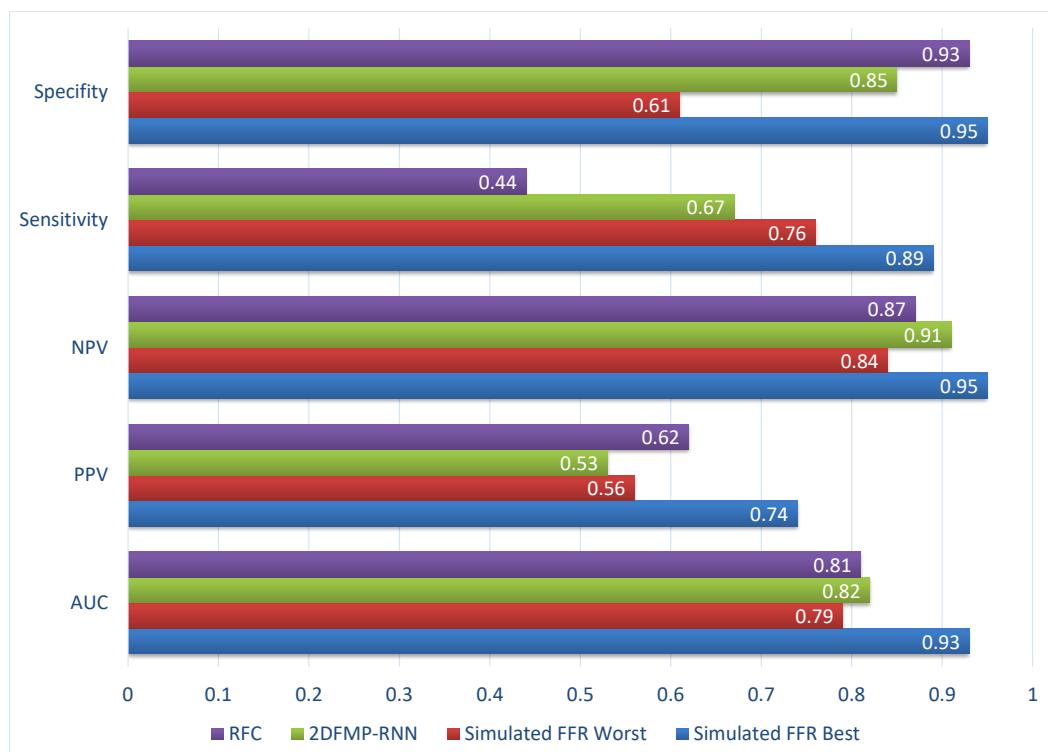


Figure 4.2: Graph of the results of our best performing methods regarding the need for revascularisation compared to simulated FFR.

Chapter 5

Summary and Outlook

5.1 Discussion

In the scope of this thesis we proposed two different approaches capable of automated high risk plaque features and stenosis significance extraction. This automatic extraction is so far not evaluated in any other publication to our knowledge, but is able to lead to a speed-up in the clinical work-flow, in the form of a CAD program. The first approach uses segmentations of the plaque segment and lumen incorporating prior information to the subsequent classification task, with the cost of neglecting some potential important information. Furthermore, radiomics introduce a new problem due to the large amount of features. These different features need to be analysed and selected with respect to their significance to the task at hand to enable the best possible classification. With our approach the amount of redundant features seemed to be high, which is a subject of possible follow up studies.

Contrary to the radiomics approach, deep learning methods only rely on the centerline extraction. Even though this would result in a speed up in the clinical workflow – since segmentation is often time consuming – segmentation will most of the time still be performed for visualisation reasons as well as to confirm the classifier’s predictions. Still errors introduced by inaccurate segmentations are not incorporated in the classification pipeline. Moreover, neural network approaches tend to enable a massive speed-up during prediction. Furthermore, all image information inside of the FOV around the centerline are used, increasing the amount of used information.

In general, one limitation of this thesis was related to the dataset used. The dataset was labeled on site by a single non-physician. Even though we based all labels on measurements an additional expert reader would help to validate the results of this thesis. Also the reference standard FFR value is not provided, making the comparison to related work hard but not impossible.

Furthermore, we had a severe class imbalance of 20% to 80% for our main target – the segment level revascularisation need –, which does impact the performance of some evaluation metrics as well as the possibility of making representative sub-sets of data during cross-validation. These sub-sets are especially important for neural networks, since they only learn what they have seen at least once during the training process. Therefore, if a prototype of segment was only in one part of the data it is impossible to predict in the other set. This can also partly not be overcome by augmenting the data with different rotations and translations, but gives us the outlook of all methods performing better for a bigger dataset.

Concluding, our results cannot provide a direct comparison to other state of the art methods, but does bear a proof of concept with evaluation metric scores close to the reference approaches.

5.2 Future Work

As previously discussed, the data collection used in the scope of this thesis was suboptimal. The methods implemented in this thesis therefore shall be applied to a different data set with given reference FFR values in order to make it comparable to simulated FFR. Also a second reader shall annotate the data from this data set in order to validate the results of this study.

There is also still a lot of room for future developments for each approach investigated. The neural network based approach can be improved by incorporating transfer learning. For a transfer learning approach the network is trained for multiple targets at once, which does make sense for our problem at hand, since most features extracted from the CNN should carry information for multiple targets. This fixed feature extraction method could then be incorporated to a prototype program in order to create completely anonymised data for a feedback based reinforcement learning algorithm. This would fix the problem of having too little data and would create a system, which is based annotations from multiple readers.

Moreover, the data used in the scope of this thesis will be used for some other problems related to coronary artery plaque: detection and segmentation of segments of interest. A robust segmentation algorithm implemented using deep learning techniques would provide a massive speed-up during pre-processing and detecting segments of interest is of high clinical value. First experiments in these directions were already performed with promising results.

The radiomics based approach could benefit from a different feature selection method, since the used method does not suppress redundant features, and therefore will be evaluated with different feature selection methods. Also in this thesis we investigated three different masks. Since the shape based radiomics correspond to altered information for each, multiple masks shall be used

during the extraction of the radiomics features and the combined information shall be evaluated.

5.3 Summary

CVDs remain the most frequent cause of death across all industrialised countries [Men15], with many of them being related to atherosclerotic plaque segments in the coronaries [Nag03]. Concluding, the interest in identifying plaque segments, which bear risk to the patient is very high. Atherosclerotic plaque segments are defined as antisymmetric focal narrowings of the lumen and are mostly caused by inflammatory reactions inside the vessel wall. Two kinds of health risking plaques exists: vulnerable plaques, which are prone to rupture, and plaque segments, which significantly narrow the lumen creating a functionally significant stenosis. The first leads to thrombus formation resulting in cardiac infarction or stroke, while the second type leads to a poorly perfused heart muscle and therefore to cardiac ischemia.

A modality usable to assess these high risk plaques non-invasively is CCTA, where a CT image of the heart is taken after injecting a contrast agent. The reference standard to interpret the resulting images is still visual assessment, which has disadvantages such as inter-/intra-viewer variability and human error. An alternate approach is simulated FFR, which tries to approximate the FFR value — defined as the ratio of intravascular pressure before and after the segment — from CCTA scans. This ratio is typically measured in an interventional setup and is the reference standard to identify functionally significant stenosis. Furthermore, so called high-risk plaque features can be identified in CCTA data, namely positive remodeling, high stenosis degree ($> 50\%$), spotty calcifications and the NRS. In the scope of this thesis we want to identify both, vulnerable plaques and functionally significant stenosis using radiomics and deep learning and integrate the resulting machine learning setups to a coronary plaque analysis program.

Radiomics contain multiple quantitative measures of pathologic characteristics from radiologic data, which may not necessarily be perceivable to the human reader. In order to calculate these metrics a prior segmentation mask of the lesion is needed and all metrics are calculated on the masked version of the original image. Radiomic features can be divided into four groups: intensity-, texture-, shape- and transformation-based metrics. Intensity-based metrics can be directly calculated from a histogram and describe the intensity distribution within the area of interest. They are already implemented in the clinical work flow but are scanner dependent. The next group — texture-based metrics — use second- and higher order statistics to evaluate the texture of the lesion. These metrics include i.e. homogeneity or contrast and are able to extract information about the composition of the plaque. Shape-based metrics mainly use the mask

information in order to extract characteristics in the spatial domain. They are well established in the clinic, but since radiomics originate from tumor analysis they rather analyse spherical objects and should be extended to cylindrical metrics for our use case. And as the last feature group transformation-based metrics transform the masked image in a different domain (e.g. frequency domain) or apply a function in order to extract different types of information from the image. However, they are partly hard to interpret.

Feature vectors resulting from the radiomic feature extraction are too high dimensional in order to be used for a standard machine learning setup. To overcome this problem feature selection is performed using the ANOVA f-score. This method selects the best features by calculating the mean and variance of the class distributions linked to each single feature. The features with the highest difference in mean and variance for the different classes are chosen since they are proven to have discriminating capabilities.

In order to classify using the resulting feature vectors four different classifiers are used: GNB assumes feature independency and uses this assumption in order to fit a gaussian distribution for each class and creating a decision boundary between them, SVMs draw a decision boundary between the classes with the maximal distance to both, random forests use an ensemble of decision trees trained on random subsets of samples and features and XGBoost also uses an ensemble of decision trees but adds new decision trees by using a greedy search to find a decision tree which minimizes a loss function. In order to increase the capacity of SVMs they are boosted with the AdaBoost algorithm, which builds an ensemble of classifiers by weighting all previously classified samples when adding a new classifier.

Deep learning refers to the training of neural networks. These neural networks utilise a chain of trainable and non-trainable functions in order to perform classification tasks. Trainable functions are called layers and mostly are dot-products with a trainable weight matrix (fully-connected layer) or convolutions with trainable kernel weights (convolutional layer). The weights of these layers get adapted by calculating the error with respect to the networks prediction and updating all layers using the resulting gradient with respect to the weights utilising the chain rule. A third form of trainable layer, the RNN cell can be used in order to analyse sequences.

In the scope of this thesis a data collection of 105 patients with 463 lesions was used. For each patient a label with respect to the revascularisation of all main coronary branches existed. They were manually relabeled with respect to their need for stenosis grade, revascularisation (based on the stenosis grade and the branch label), low HU attenuation, spotty calcifications and positive remodeling. Due to inaccurate annotations, irrelevant segments, lesions at bifurcations and heavy noise several segments needed to be excluded.

For the radiomics based approach three different sorts of masks were used and evaluated for the different targets: lumen only, plaque only and both combined. The evaluation was performed in a 10-fold cross validation manner using features selected using the training set only with the ANOVA f-score and classifying with the four above mentioned classifiers. In order to enable the deep learning approach the data needed to be adapted. Therefore, MPR image stacks, which were interpolated orthogonal to the vessel centerline were extracted for each segment. Since the training of neural networks is more stable when all inputs have the same size, they were padded with zeros or resized to the same size. Four different models were used for the final evaluation, differing in the kind of feature extraction and sequence analysis. Each model was trained ten times for 50 epochs in a cross validation manner for each target. In order to assess the performance of both approaches the AUC, accuracy, F1 score, sensitivity, specificity, PPV, NPV and MCC were calculated.

Both approaches performed well on most targets, but the positive remodeling, where the best radiomics-based classifier only scored an AUC of 0.66 and an MCC of 0.27, while the best deep learning model reached an AUC of 0.69 and an MCC of 0.26. Other than that, all other targets on the lesion level performed very good with an AUC greater than 0.8 and an MCC of at least 0.42. Especially the stenosis grade and the low HU value were classified with an AUC of greater than 0.9 and an MCC of at least 0.58. Within each approach the different models and classifiers performed very similarly. However, the kind of features used — which is determined by the kind of mask used for the radiomics approach and how the image stack was padded for the deep learning approach — prove to be of high significance for different targets. The most important target — the need for revascularisation — was compared to results of the simulated FFR measurements. Even though the labels used for our study were not based on the FFR value during labeling, it is likely that the revascularisation decision was due to FFR values of the segments labeled, making them at least comparable but not direct comparable. For this target several different studies scored an AUC, PPV, NPV, sensitivity and specificity between 0.79-0.93, 0.56-0.74, 0.84-0.95, 0.76-0.89 and 0.61-0.95, while our deep learning approach scored 0.82, 0.53, 0.91, 0.67 and 0.85 and our radiomics approach scored 0.81, 0.62, 0.87, 0.44 and 0.93. Therefore our metrics lie mostly within the range of these other studies, which is a strong hint towards the validity of our approaches.

5.4 Conclusion

We presented and evaluated two different approaches for the automated extraction of high risk plaque features from CCTA scans. They are based on radiomics and deep learning and implement state of the art machine learning techniques. Both methods were evaluated using 10-fold cross validation. They performed well for most targets and show comparable metric scores to state of the art approaches evaluated on different data-sets. Also they were partly integrated to a coronary artery plaque analysis program, validating their applicability in the clinical workflow. By integrating such an automated tool that extracts high risk plaque information, automated reports could be suggested by a computer assisted diagnosis program, which speeds up the work flow and can serve as a second opinion for physicians.

Acronyms

ACS adverse coronary syndrom

Adam adaptive moment estimation

AdaBoost adaptive boosting

AI artificial intelligence

ANOVA analysis of variance

AUC area under curve

CAD-RADS coronary artery disease reporting and data system

CAD computer assisted diagnosis

CCB Cardioangiologischem Centrum Bethanien

CCTA coronary computed tomography angiography

CT computed tomography

CNN convolutional neural network

CPR curved planar reformation

CVD Cardiovascular disease

CX circumflex artery

ESS endothelial shear stress

FFR fractional flow reserve

FMP fractional max-pooling

FOV field of view

GLCM gray level co-occurrence matrix

GLRLM gray level run length matrix

GLSZM gray level size zone matrix

GNB gaussian naïve Bayes

GRU gated recurrent unit

GUI graphical user interface

MCC Matthews correlation coefficient

MLP multi-layer perceptron

NGTDM neighbourhood gray-tone difference matrix

NPV negative predictive value

GLDM gray level dependence matrix

HU Hounsfield unit

TCFA thin-cap fibroatheroma

IOCT intravascular optical coherence tomography

IVUS intravascular ultrasound

LSTM long short-term memory unit

LAD left anterior descending artery

MPR multi planar reformatted

NRS napkin-ring sign

PPV positive predictive value

RBF radial basis function

RCA right coronary artery

RCNN recurrent convolutional neural network

ResNet residual network

ReLU rectified linear unit

RFC random forest classifier

RNN recurrent neural network

ROI region of interest

SVM support vector machine

XGBoost extreme gradient boosting

Appendix A

Data Analysis

	Stenosis	Positive Remodeling	Spotty Calcifications	low HU attenuation
No Revascularisation	8	236	203	16
Revascularisation	72	48	26	16
Total	80	284	229	32

Table A.1: Co-occurrence of the need of revascularisation and other targets

	Revascularisation	Positive Remodeling	Spotty Calcifications	low HU attenuation
No Stenosis	21	245	218	16
Stenosis	72	39	11	16
Total	93	284	229	32

Table A.2: Co-occurrence of the stenosis grade and other targets

	Positive Remodeling	Spotty Calcifications	low HU attenuation
No Stenosis/No Revascularisation	232	202	15
No Stenosis/Revascularisation	4	1	1
Stenosis/No Revascularisation	13	16	1
Stenosis/Revascularisation	35	10	15
Total	284	229	32

Table A.3: Co-occurrence of stenosis grade with revascularisation against other targets

	Revascularisation	Stenosis	Positive Remodeling	Low HU Attenuation
No Spotty Calcifications	67	69	116	28
Spotty Calcifications	26	11	168	4
Total	93	80	284	32

Table A.4: Co-occurrence of spotty calcifications and other targets

	Revascularisation	Stenosis	Positive Remodeling	Spotty Calcifications
No Stenosis	77	64	268	28
Stenosis	16	16	16	4
Total	93	80	284	229

Table A.5: Co-occurrence of low HU attenuation and other targets

Appendix B

Models

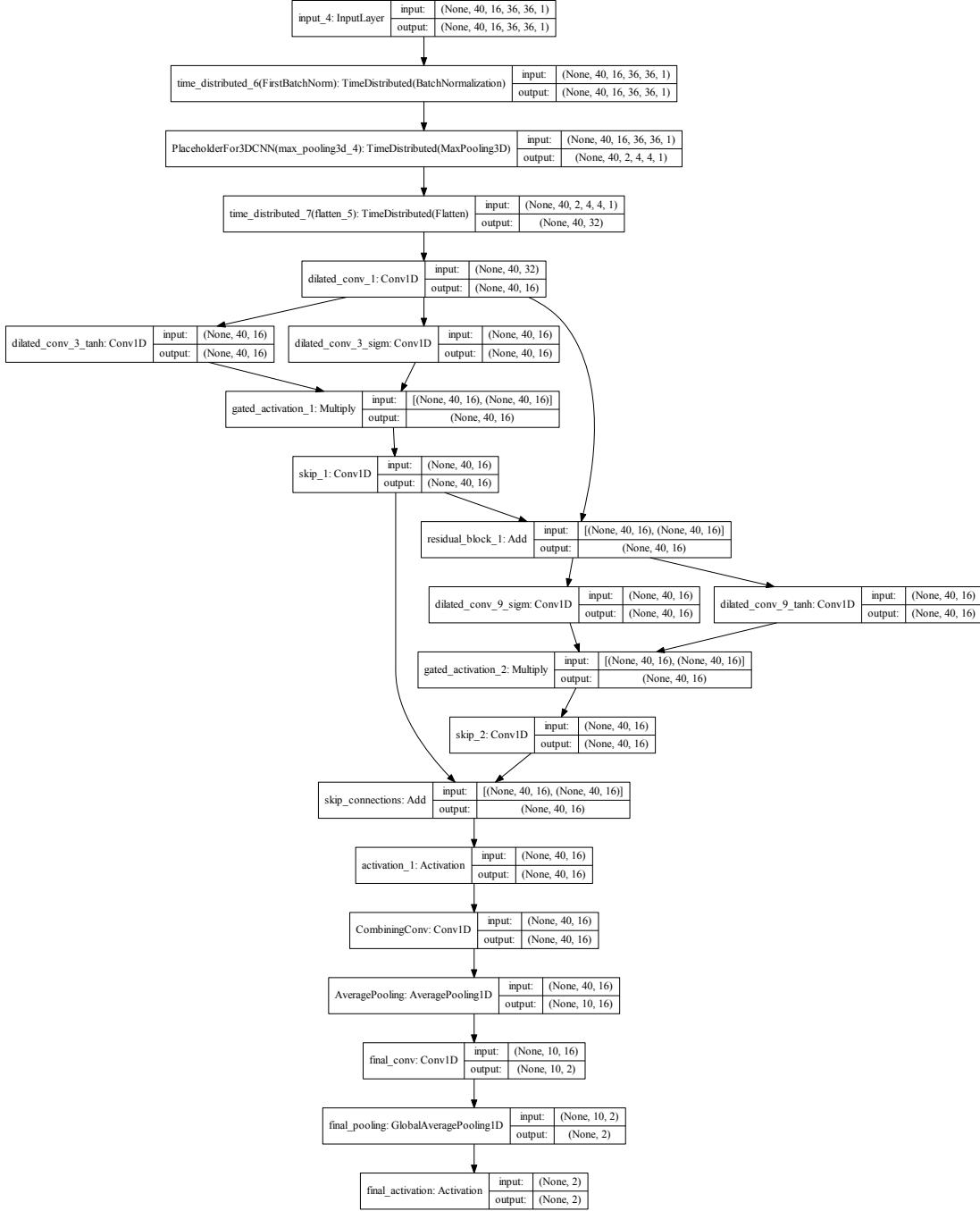


Figure B.1: 3D convolutions for feature extraction with WaveNet sequence analysis; Note that for this representation the feature extraction CNN is replaced by a MaxPooling Layer – the CNN architecture is similar to the one of the 3DRCNN model– and that a dilation rate of 2 was used in contrast to the dilation rate of 8 used in the scope of this thesis.

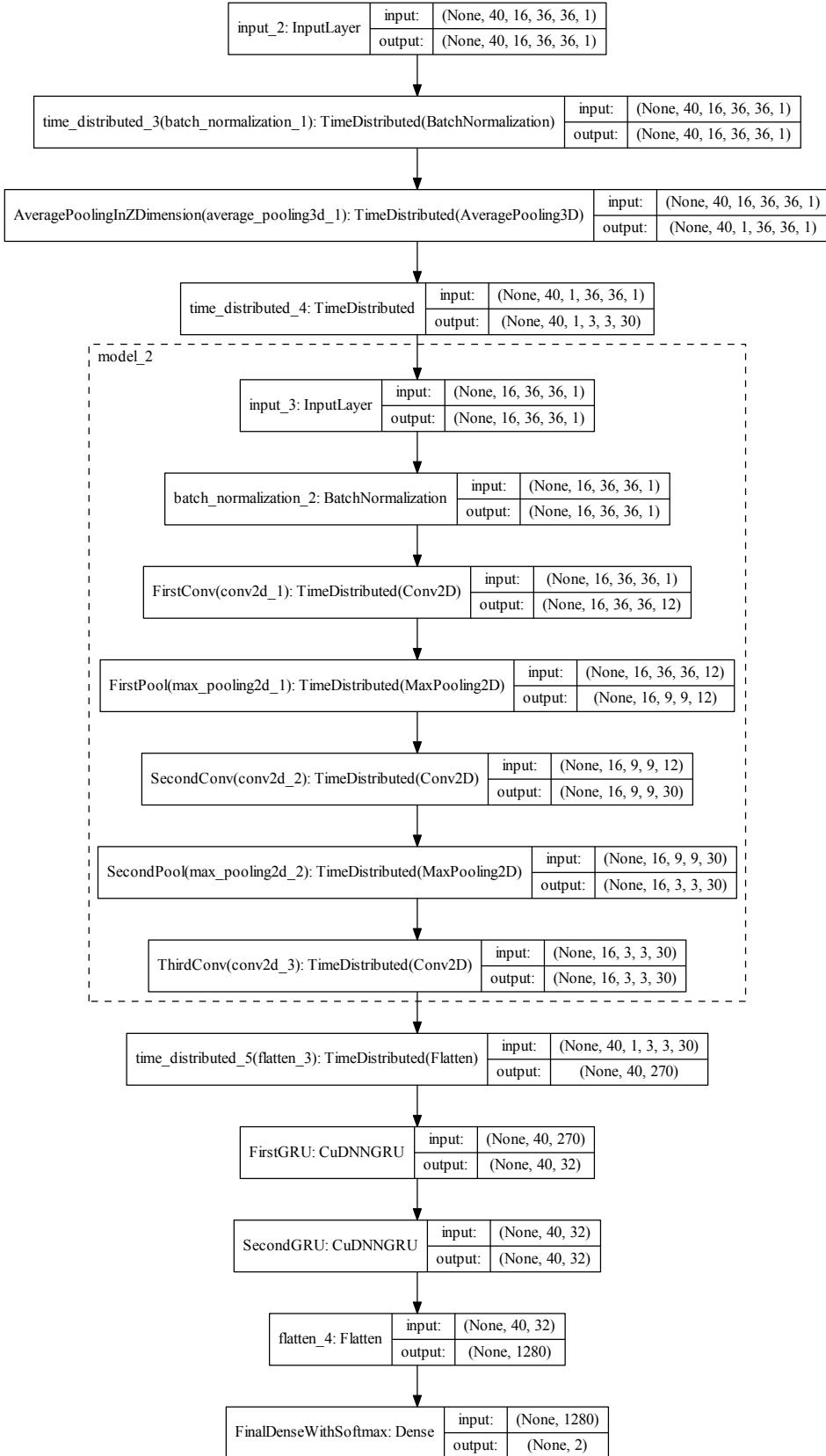


Figure B.2: 2D convolutions for feature extraction with RNN sequence analysis

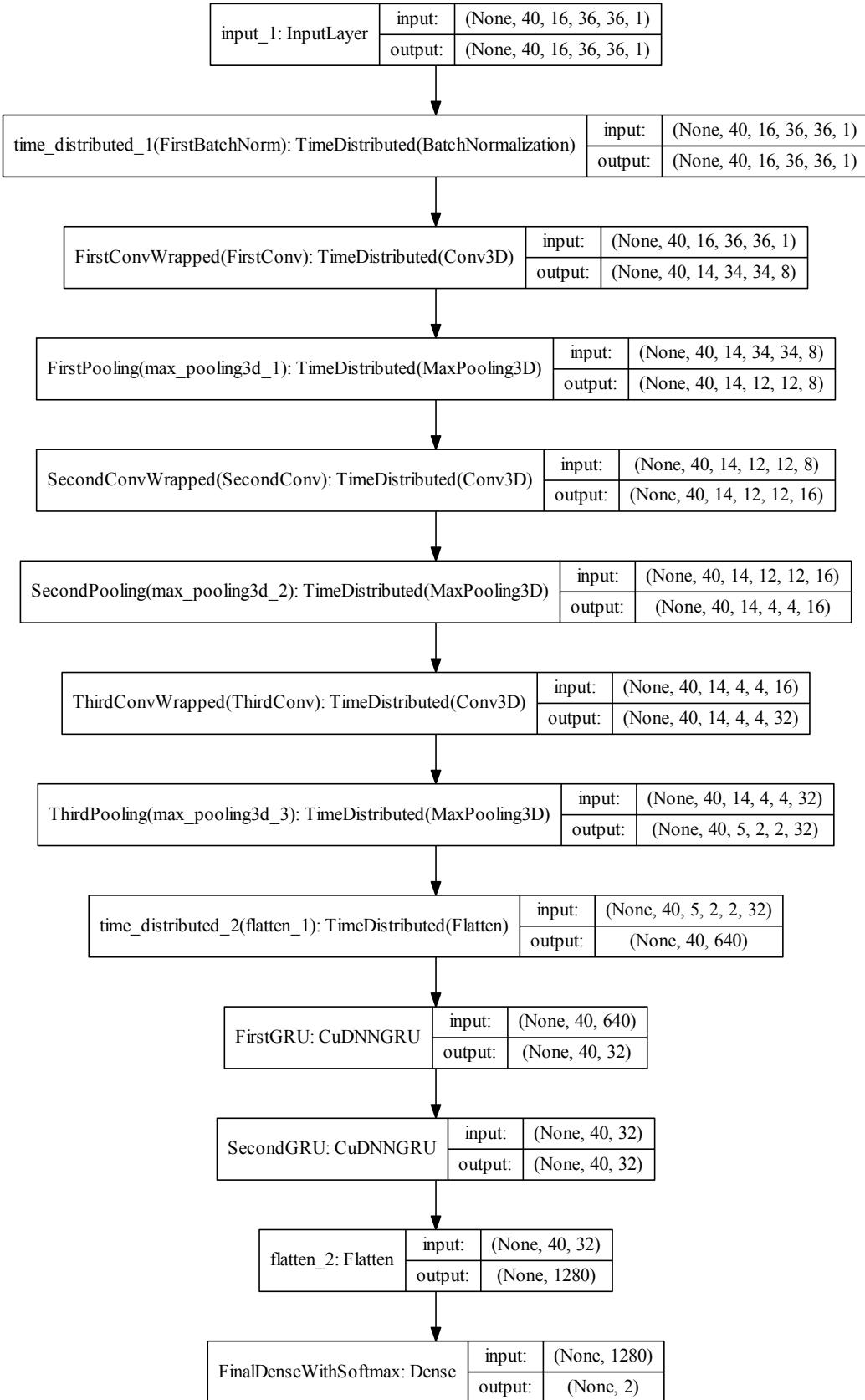


Figure B.3: 3D convolutions for feature extraction with RNN sequence analysis

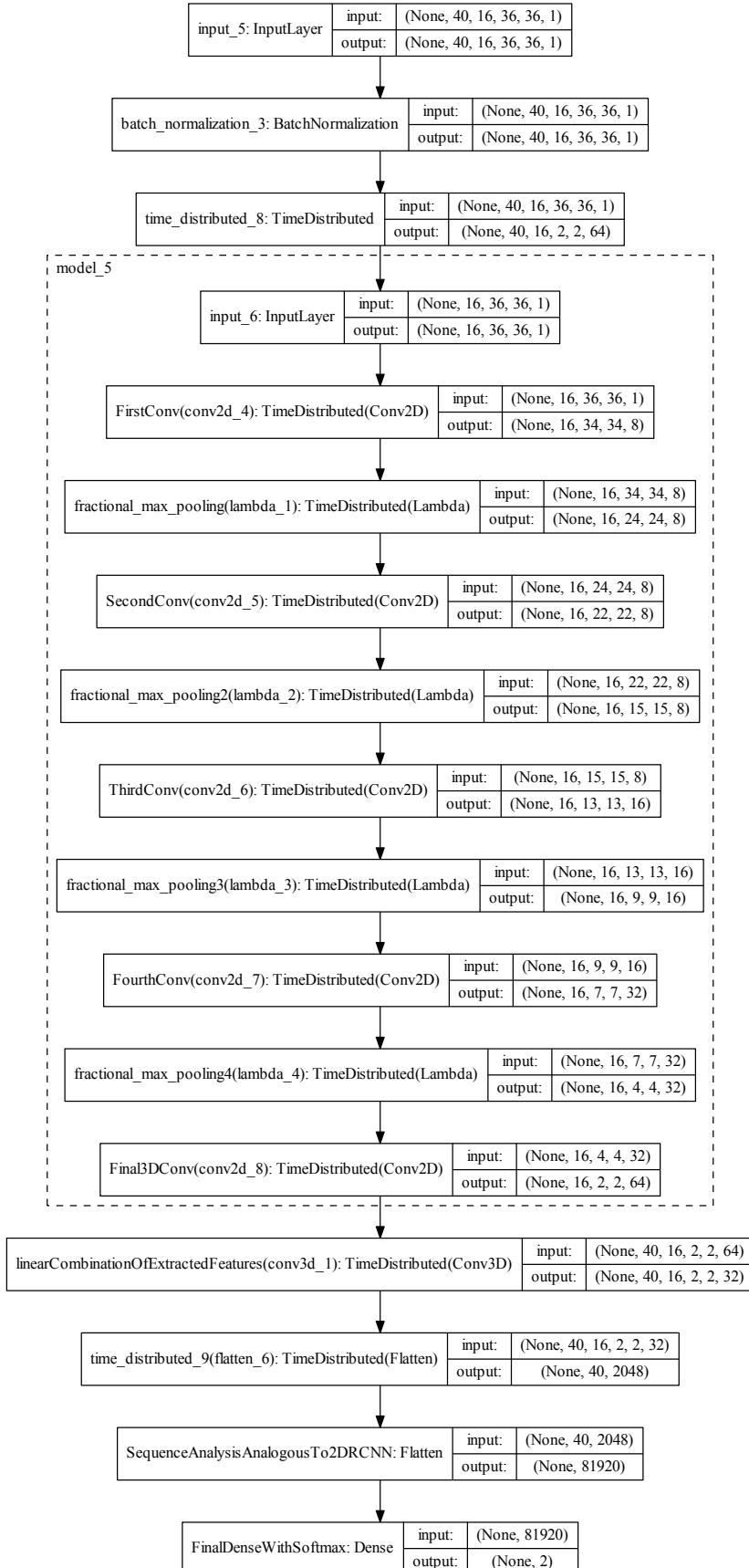


Figure B.4: 2D convolutions with fractional max pooling for feature extraction and RNN sequence analysis

Appendix C

Detailed Results

Best Results

Table C.1: Best Results

Target	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Stenosis > 50%	RFC	0.91	0.85	0.68	0.74	0.88	0.64	0.92	0.59
Stenosis > 50%	3DWave	0.89	0.87	0.66	0.64	0.93	0.69	0.91	0.58
Remodeling	RFC	0.66	0.67	0.76	0.68	0.63	0.86	0.37	0.27
Remodeling	3DRNN	0.69	0.65	0.72	0.72	0.54	0.72	0.55	0.26
Low HU	GNB	0.96	0.94	0.67	0.55	0.99	0.84	0.95	0.65
Low HU	3DRNN	0.89	0.91	0.49	0.38	0.98	0.72	0.92	0.48
Spotty Calc	GNB	0.86	0.84	0.84	0.84	0.85	0.84	0.84	0.69
Spotty Calc	3DRNN	0.86	0.82	0.82	0.84	0.81	0.80	0.85	0.65
Revasc Plaque	RFC	0.81	0.83	0.52	0.62	0.87	0.44	0.93	0.42
Revasc Plaque	2DFMP	0.82	0.81	0.59	0.53	0.91	0.67	0.85	0.47
Revasc Branch	adaSVM	0.75	0.80	0.47	0.68	0.82	0.36	0.95	0.39
Revasc Patient	GNB	0.79	0.71	0.65	0.72	0.70	0.59	0.81	0.41

C.1 Radiomics Mask Comparison

Table C.2: Stenosis Degree

Mask	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Lumen only	RFC	0.91	0.85	0.68	0.74	0.88	0.64	0.92	0.59
Lumen only	GNB	0.88	0.79	0.66	0.54	0.93	0.83	0.77	0.53
Lumen only	ADASVM	0.92	0.84	0.65	0.72	0.88	0.60	0.92	0.56
Lumen only	XGBoost	0.90	0.84	0.67	0.67	0.89	0.66	0.89	0.56
Lumen + Plaque	RFC	0.84	0.81	0.57	0.67	0.84	0.49	0.92	0.46
Lumen + Plaque	GNB	0.86	0.79	0.64	0.55	0.91	0.75	0.80	0.50
Lumen + Plaque	ADASVM	0.86	0.82	0.58	0.69	0.85	0.50	0.92	0.48
Lumen + Plaque	XGBoost	0.85	0.83	0.62	0.71	0.86	0.56	0.92	0.53
Plaque only	RFC	0.82	0.79	0.53	0.61	0.84	0.47	0.90	0.41
Plaque only	GNB	0.79	0.56	0.58	0.85	0.55	0.87	0.42	0.31
Plaque only	ADASVM	0.82	0.80	0.54	0.63	0.84	0.47	0.91	0.42
Plaque only	XGBoost	0.81	0.78	0.54	0.57	0.84	0.51	0.87	0.40

Table C.3: Positive Remodeling

Mask	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Lumen only	RFC	0.52	0.57	0.68	0.63	0.44	0.74	0.32	0.06
Lumen only	GNB	0.55	0.59	0.70	0.63	0.46	0.77	0.30	0.09
Lumen only	ADASVM	0.55	0.58	0.71	0.62	0.43	0.82	0.21	0.04
Lumen only	XGBoost	0.51	0.56	0.66	0.63	0.42	0.69	0.35	0.04
Lumen + Plaque	RFC	0.57	0.62	0.72	0.65	0.52	0.80	0.33	0.15
Lumen + Plaque	GNB	0.58	0.61	0.70	0.66	0.50	0.75	0.38	0.14
Lumen + Plaque	ADASVM	0.57	0.60	0.71	0.64	0.48	0.81	0.28	0.10
Lumen + Plaque	XGBoost	0.54	0.60	0.68	0.66	0.48	0.70	0.43	0.14
Plaque only	RFC	0.66	0.67	0.76	0.68	0.63	0.86	0.37	0.27
Plaque only	GNB	0.58	0.69	0.79	0.68	0.74	0.93	0.31	0.32
Plaque only	ADASVM	0.64	0.66	0.75	0.69	0.59	0.82	0.41	0.26
Plaque only	XGBoost	0.63	0.62	0.71	0.67	0.52	0.76	0.41	0.18

Table C.4: Low HU Attenuation

Mask	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Lumen only	RFC	0.67	0.93	0.00	0.00	0.93	0.00	1,00	-0.01
Lumen only	GNB	0.61	0.87	0.15	0.14	0.94	0.16	0.92	0.08
Lumen only	ADASVM	0.62	0.91	0.11	0.20	0.93	0.08	0.98	0.09
Lumen only	XGBoost	0.66	0.91	0.06	0.13	0.93	0.04	0.98	0.03
Lumen + Plaque	RFC	0.89	0.95	0.54	0.69	0.96	0.44	0.98	0.52
Lumen + Plaque	GNB	0.89	0.93	0.60	0.51	0.98	0.72	0.95	0.57
Lumen + Plaque	ADASVM	0.91	0.93	0.50	0.48	0.96	0.52	0.96	0.46
Lumen + Plaque	XGBoost	0.91	0.94	0.52	0.65	0.96	0.44	0.98	0.51
Plaque only	RFC	0.94	0.95	0.61	0.67	0.97	0.56	0.98	0.58
Plaque only	GNB	0.96	0.94	0.67	0.55	0.99	0.84	0.95	0.65
Plaque only	ADASVM	0.96	0.93	0.50	0.52	0.96	0.48	0.97	0.46
Plaque only	XGBoost	0.92	0.93	0.45	0.53	0.96	0.40	0.97	0.42

Table C.5: Spotty Calcification

Mask	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Lumen only	RFC	0.85	0.82	0.81	0.82	0.81	0.80	0.83	0.63
Lumen only	GNB	0.86	0.84	0.84	0.84	0.85	0.84	0.84	0.69
Lumen only	ADASVM	0.86	0.84	0.84	0.85	0.84	0.83	0.86	0.69
Lumen only	XGBoost	0.84	0.82	0.81	0.83	0.80	0.79	0.84	0.63
Lumen + Plaque	RFC	0.84	0.81	0.81	0.82	0.81	0.79	0.83	0.63
Lumen + Plaque	GNB	0.86	0.82	0.82	0.82	0.83	0.82	0.83	0.65
Lumen + Plaque	ADASVM	0.85	0.83	0.82	0.85	0.82	0.80	0.87	0.67
Lumen + Plaque	XGBoost	0.84	0.80	0.79	0.81	0.78	0.76	0.83	0.59
Plaque only	RFC	0.87	0.82	0.81	0.83	0.81	0.79	0.84	0.64
Plaque only	GNB	0.86	0.82	0.81	0.83	0.80	0.78	0.85	0.63
Plaque only	ADASVM	0.87	0.83	0.83	0.83	0.83	0.83	0.83	0.66
Plaque only	XGBoost	0.85	0.82	0.81	0.83	0.81	0.79	0.84	0.63

Table C.6: Revascularisation Plaque Level

Mask	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Lumen only	RFC	0.81	0.83	0.52	0.62	0.87	0.44	0.93	0.42
Lumen only	GNB	0.80	0.76	0.55	0.44	0.91	0.71	0.77	0.41
Lumen only	ADASVM	0.78	0.81	0.47	0.56	0.86	0.40	0.92	0.37
Lumen only	XGBoost	0.81	0.81	0.45	0.53	0.85	0.39	0.91	0.34
Lumen + Plaque	RFC	0.77	0.81	0.41	0.58	0.84	0.32	0.94	0.33
Lumen + Plaque	GNB	0.78	0.77	0.51	0.44	0.89	0.61	0.80	0.37
Lumen + Plaque	ADASVM	0.72	0.80	0.36	0.51	0.83	0.28	0.93	0.27
Lumen + Plaque	XGBoost	0.79	0.83	0.50	0.61	0.86	0.43	0.93	0.41
Plaque only	RFC	0.71	0.80	0.29	0.50	0.82	0.21	0.95	0.22
Plaque only	GNB	0.70	0.78	0.45	0.46	0.86	0.44	0.87	0.32
Plaque only	ADASVM	0.72	0.81	0.32	0.59	0.83	0.22	0.96	0.28
Plaque only	XGBoost	0.71	0.78	0.30	0.41	0.82	0.24	0.91	0.19

Table C.7: Revascularisation Plaque Level Edgecases

Mask	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Lumen only	RFC	0.54	0.78	0.21	0.25	0.85	0.18	0.90	0.09
Lumen only	GNB	0.39	0.54	0.11	0.08	0.80	0.18	0.60	-0.16
Lumen only	ADASVM	0.50	0.84	0.00	nan	0.84	0.00	1,00	0.00
Lumen only	XGBoost	0.46	0.71	0.09	0.09	0.83	0.09	0.83	-0.08
Lumen + Plaque	RFC	0.49	0.75	0.11	0.13	0.84	0.09	0.88	-0.03
Lumen + Plaque	GNB	0.39	0.54	0.11	0.08	0.80	0.18	0.60	-0.16
Lumen + Plaque	ADASVM	0.50	0.84	0.00	nan	0.84	0.00	1,00	0.00
Lumen + Plaque	XGBoost	0.59	0.81	0.32	0.38	0.87	0.27	0.91	0.21
Plaque only	RFC	0.54	0.78	0.21	0.25	0.85	0.18	0.90	0.09
Plaque only	GNB	0.50	0.71	0.17	0.15	0.84	0.18	0.81	-0.01
Plaque only	ADASVM	0.50	0.84	0.00	nan	0.84	0.00	1,00	0.00
Plaque only	XGBoost	0.55	0.80	0.22	0.29	0.85	0.18	0.91	0.12

Table C.8: Revascularisation Branch Level

Mask	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Lumen only	RFC	0.74	0.76	0.41	0.52	0.81	0.33	0.90	0.28
Lumen only	GNB	0.72	0.69	0.39	0.38	0.80	0.40	0.78	0.19
Lumen only	ADASVM	0.67	0.74	0.35	0.44	0.79	0.29	0.88	0.20
Lumen only	XGBoost	0.72	0.74	0.41	0.47	0.81	0.36	0.87	0.25
Lumen + Plaque	RFC	0.75	0.78	0.41	0.59	0.81	0.31	0.93	0.31
Lumen + Plaque	GNB	0.72	0.78	0.46	0.57	0.82	0.38	0.91	0.34
Lumen + Plaque	ADASVM	0.75	0.80	0.47	0.68	0.82	0.36	0.95	0.39
Lumen + Plaque	XGBoost	0.72	0.78	0.49	0.56	0.83	0.43	0.89	0.35
Plaque only	RFC	0.65	0.76	0.32	0.50	0.79	0.24	0.92	0.22
Plaque only	GNB	0.68	0.77	0.40	0.57	0.81	0.31	0.92	0.29
Plaque only	ADASVM	0.67	0.74	0.33	0.46	0.79	0.26	0.90	0.20
Plaque only	XGBoost	0.60	0.74	0.31	0.43	0.79	0.24	0.90	0.17

Table C.9: Revascularisation Patient Level

Mask	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Lumen only	RFC	0.78	0.70	0.67	0.67	0.72	0.67	0.72	0.39
Lumen only	GNB	0.79	0.71	0.65	0.72	0.70	0.59	0.81	0.41
Lumen only	ADASVM	0.77	0.71	0.68	0.68	0.73	0.67	0.74	0.41
Lumen only	XGBoost	0.74	0.70	0.67	0.67	0.72	0.67	0.72	0.39
Lumen + Plaque	RFC	0.72	0.69	0.62	0.69	0.69	0.56	0.79	0.36
Lumen + Plaque	GNB	0.76	0.66	0.58	0.67	0.66	0.51	0.79	0.31
Lumen + Plaque	ADASVM	0.72	0.71	0.68	0.68	0.73	0.67	0.74	0.41
Lumen + Plaque	XGBoost	0.69	0.70	0.65	0.69	0.71	0.62	0.77	0.39
Plaque only	RFC	0.73	0.62	0.56	0.58	0.64	0.54	0.68	0.22
Plaque only	GNB	0.72	0.67	0.59	0.69	0.67	0.51	0.81	0.34
Plaque only	ADASVM	0.67	0.67	0.63	0.65	0.69	0.62	0.72	0.34
Plaque only	XGBoost	0.73	0.70	0.66	0.68	0.71	0.64	0.74	0.39

C.2 Selected Features

Table C.10: Selected Features 1

Mask	Target	Image Transformation	Feature Group	Feature
Lumen only	Stenosis	original	shape	Compactness1
Lumen only	Stenosis	original	shape	Compactness2
Lumen only	Stenosis	original	shape	Elongation
Lumen only	Stenosis	original	shape	Flatness
Lumen only	Stenosis	original	shape	MajorAxis
Lumen only	Stenosis	original	shape	Maximum3DDiameter
Lumen only	Stenosis	original	shape	SphericalDisproportion
Lumen only	Stenosis	original	shape	Sphericity
Lumen only	Stenosis	original	shape	SurfaceArea
Lumen only	Stenosis	squareroot	glszm	GrayLevelNonUniformity
Plaque only	Remodeling	exponential	glcm	Imc2
Plaque only	Remodeling	exponential	glcm	InverseVariance
Plaque only	Remodeling	exponential	glszm	GrayLevelNonUniformity
Plaque only	Remodeling	exponential	glszm	ZoneEntropy
Plaque only	Remodeling	original	glcm	Imc2
Plaque only	Remodeling	square	glcm	Imc2
Plaque only	Remodeling	square	glszm	GrayLevelNonUniformity
Plaque only	Remodeling	squareroot	firstrorder	Kurtosis
Plaque only	Remodeling	squareroot	firstrorder	Skewness
Plaque only	Remodeling	wavelet-LLL	glcm	Imc2
Plaque only	Low HU	original	gldm	LargeDependenceEmphasis
Plaque only	Low HU	original	glrlm	LongRunEmphasis
Plaque only	Low HU	original	glrlm	RunVariance
Plaque only	Low HU	square	glcm	JointEnergy
Plaque only	Low HU	square	glcm	MaximumProbability
Plaque only	Low HU	square	gldm	DependenceVariance
Plaque only	Low HU	square	gldm	LargeDependenceEmphasis
Plaque only	Low HU	square	gldm	LargeDependenceLowGray
Plaque only	Low HU	square	glrlm	RunLengthNonUniformity
Plaque only	Low HU	square	glrlm	ShortRunEmphasis
Lumen only	Spotty	original	glszm	GrayLevelNonUniformity
Lumen only	Spotty	original	shape	Compactness1
Lumen only	Spotty	original	shape	Compactness2
Lumen only	Spotty	original	shape	Elongation
Lumen only	Spotty	original	shape	Flatness
Lumen only	Spotty	original	shape	MajorAxis
Lumen only	Spotty	original	shape	Maximum3DDiameter
Lumen only	Spotty	original	shape	SphericalDisproportion
Lumen only	Spotty	original	shape	Sphericity
Lumen only	Spotty	original	shape	SurfaceArea

Table C.11: Selected Features 2

Mask	Target	Image Transformation	FeatureGroup	Feature
Lumen only	Revasc Plaque	original	shape	Compactness1
Lumen only	Revasc Plaque	original	shape	Compactness2
Lumen only	Revasc Plaque	original	shape	Elongation
Lumen only	Revasc Plaque	original	shape	Flatness
Lumen only	Revasc Plaque	original	shape	MajorAxis
Lumen only	Revasc Plaque	original	shape	Maximum2DDiameter
Lumen only	Revasc Plaque	original	shape	Maximum3DDiameter
Lumen only	Revasc Plaque	original	shape	SphericalDisproportion
Lumen only	Revasc Plaque	original	shape	Sphericity
Lumen only	Revasc Plaque	squareroot	glcm	MaximumProbability
Lumen and plaque	Revasc Branch	logarithm	firstorder	InterquartileRange
Lumen and plaque	Revasc Branch	logarithm	firstorder	RobustMeanAbsolute
Lumen and plaque	Revasc Branch	original	shape	Compactness1
Lumen and plaque	Revasc Branch	original	shape	Compactness2
Lumen and plaque	Revasc Branch	original	shape	Maximum3DDiameter
Lumen and plaque	Revasc Branch	original	shape	SphericalDisproportion
Lumen and plaque	Revasc Branch	original	shape	Sphericity
Lumen and plaque	Revasc Branch	original	firstorder	Meanvar
Lumen and plaque	Revasc Branch	original	firstorder	Medianvar
Lumen and plaque	Revasc Branch	square	gldm	LowGrayLevelEmphasis
Lumen only	Revasc Patient	original	shape	Compactness1
Lumen only	Revasc Patient	original	shape	Compactness2
Lumen only	Revasc Patient	original	shape	SphericalDisproportion
Lumen only	Revasc Patient	original	shape	Sphericity
Lumen only	Revasc Patient	wavelet-HHH	glcm	Idn
Lumen only	Revasc Patient	wavelet-HHH	glcm	Imc2
Lumen only	Revasc Patient	wavelet-HHH	glszm	LowGrayLevelZone
Lumen only	Revasc Patient	wavelet-HLH	glcm	Imc1
Lumen only	Revasc Patient	wavelet-HLH	glcm	Imc2
Lumen only	Revasc Patient	wavelet-LHL	glszm	SizeZoneNonUniformity

C.3 Radiomics Deep Learning Comparison

Table C.12: Stenosis Degree

Padding/Mask	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Lumen only	RFC	0.91	0.85	0.68	0.74	0.88	0.64	0.92	0.59
Lumen only	GNB	0.88	0.79	0.66	0.54	0.93	0.83	0.77	0.53
Lumen only	ADASVM	0.92	0.84	0.65	0.72	0.88	0.60	0.92	0.56
Lumen only	XGBoost	0.90	0.84	0.67	0.67	0.89	0.66	0.89	0.56
Resize	2D	0.82	0.83	0.56	0.56	0.90	0.57	0.90	0.46
Resize	3D	0.87	0.83	0.58	0.56	0.90	0.59	0.89	0.47
Resize	2DFMP	0.88	0.81	0.61	0.50	0.95	0.80	0.81	0.52
Resize	3DWavenet	0.89	0.87	0.66	0.64	0.93	0.69	0.91	0.58
Zeros	2D	0.87	0.79	0.60	0.47	0.95	0.84	0.78	0.51
Zeros	3D	0.87	0.80	0.58	0.48	0.93	0.73	0.81	0.47
Zeros	2DFMP	0.87	0.80	0.60	0.51	0.93	0.80	0.81	0.52
Zeros	3DWavenet	0.88	0.84	0.64	0.56	0.93	0.74	0.86	0.54

Table C.13: Positive Remodeling

Padding/Mask	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Plaque only	RFC	0.66	0.67	0.76	0.68	0.63	0.86	0.37	0.27
Plaque only	GNB	0.58	0.69	0.79	0.68	0.74	0.93	0.31	0.32
Plaque only	ADASVM	0.64	0.66	0.75	0.69	0.59	0.82	0.41	0.26
Plaque only	XGBoost	0.63	0.62	0.71	0.67	0.52	0.76	0.41	0.18
Resize	3D	0.69	0.65	0.72	0.72	0.54	0.72	0.55	0.26
Resize	2D	0.69	0.65	0.71	0.72	0.53	0.71	0.54	0.25
Resize	2DFMP	0.61	0.62	0.69	0.70	0.50	0.67	0.53	0.20
Resize	3D Wavenet	0.65	0.62	0.67	0.72	0.50	0.63	0.61	0.23
Zeros	3D	0.58	0.57	0.63	0.67	0.44	0.59	0.53	0.12
Zeros	2D	0.55	0.58	0.65	0.67	0.45	0.64	0.47	0.11
Zeros	2DFMP	0.59	0.61	0.68	0.68	0.48	0.68	0.48	0.16
Zeros	3D Wavenet	0.64	0.66	0.72	0.73	0.55	0.71	0.57	0.28

Table C.14: Low HU Attenuation

Padding/Mask	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Plaque only	RFC	0.94	0.95	0.61	0.67	0.97	0.56	0.98	0.58
Plaque only	GNB	0.96	0.94	0.67	0.55	0.99	0.84	0.95	0.65
Plaque only	ADASVM	0.96	0.93	0.50	0.52	0.96	0.48	0.97	0.46
Plaque only	XGBoost	0.92	0.93	0.45	0.53	0.96	0.40	0.97	0.42
Resize	3D	0.89	0.91	0.49	0.38	0.98	0.72	0.92	0.48
Resize	2D	0.86	0.90	0.47	0.35	0.98	0.68	0.92	0.44
Resize	2DFMP	0.88	0.86	0.36	0.26	0.97	0.60	0.88	0.33
Resize	3DWavenet	0.88	0.85	0.35	0.25	0.96	0.64	0.88	0.40
Zeros	3D	0.79	0.76	0.29	0.18	0.98	0.76	0.76	0.29
Zeros	2D	0.75	0.74	0.26	0.16	0.97	0.72	0.74	0.25
Zeros	2DFMP	0.79	0.88	0.29	0.24	0.95	0.36	0.92	0.23
Zeros	3DWavenet	0.81	0.85	0.31	0.22	0.96	0.52	0.88	0.27

Table C.15: Spotty Calcification

Padding/Mask	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Lumen only	RFC	0.85	0.82	0.81	0.82	0.81	0.80	0.83	0.63
Lumen only	GNB	0.86	0.84	0.84	0.84	0.85	0.84	0.84	0.69
Lumen only	ADASVM	0.86	0.84	0.84	0.85	0.84	0.83	0.86	0.69
Lumen only	XGBoost	0.84	0.82	0.81	0.83	0.80	0.79	0.84	0.63
Resize	3D	0.85	0.80	0.79	0.82	0.78	0.77	0.83	0.60
Resize	2D	0.85	0.81	0.80	0.81	0.80	0.80	0.81	0.61
Resize	2DFMP	0.84	0.79	0.78	0.81	0.77	0.74	0.83	0.58
Resize	3D Wavenet	0.82	0.75	0.77	0.71	0.81	0.84	0.66	0.51
Zeros	3D	0.86	0.82	0.82	0.84	0.81	0.80	0.85	0.65
Zeros	2D	0.86	0.82	0.82	0.82	0.82	0.81	0.82	0.64
Zeros	2DFMP	0.86	0.82	0.81	0.81	0.82	0.81	0.82	0.63
Zeros	3D Wavenet	0.86	0.82	0.81	0.83	0.81	0.80	0.84	0.64

Table C.16: Revascularisation Plaque Level

Padding/Mask	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Lumen only	RFC	0.81	0.83	0.52	0.62	0.87	0.44	0.93	0.42
Lumen only	GNB	0.80	0.76	0.55	0.44	0.91	0.71	0.77	0.41
Lumen only	ADASVM	0.78	0.81	0.47	0.56	0.86	0.40	0.92	0.37
Lumen only	XGBoost	0.81	0.81	0.45	0.53	0.85	0.39	0.91	0.34
Resize	3D	0.78	0.76	0.49	0.43	0.88	0.57	0.81	0.34
Resize	2D	0.77	0.75	0.52	0.43	0.90	0.65	0.77	0.37
Resize	2DFMP	0.82	0.76	0.57	0.46	0.92	0.75	0.77	0.45
Resize	3DWavenet	0.77	0.75	0.49	0.43	0.88	0.58	0.80	0.34
Zeros	3D	0.79	0.76	0.53	0.45	0.90	0.67	0.78	0.39
Zeros	2D	0.78	0.75	0.53	0.43	0.90	0.69	0.76	0.39
Zeros	2DFMP	0.82	0.81	0.59	0.53	0.91	0.67	0.85	0.47
Zeros	3DWavenet	0.77	0.75	0.52	0.44	0.89	0.64	0.78	0.38

Table C.17: Revascularisation Plaque Level Edgecases

Padding/Mask	Classifier	AUC	ACC	F1	PPV	NPV	Sensitivity	Specificity	MCC
Lumen only	RFC	0.54	0.78	0.21	0.25	0.85	0.18	0.90	0.09
Lumen only	GNB	0.39	0.54	0.11	0.08	0.80	0.18	0.60	-0.16
Lumen only	ADASVM	0.50	0.84	0.00	nan	0.84	0.00	1,00	0.00
Lumen only	XGBoost	0.46	0.71	0.09	0.09	0.83	0.09	0.83	-0.08
Resize	3D	0.51	0.53	0.28	0.20	0.81	0.47	0.55	0.01
Resize	2D	0.50	0.56	0.29	0.21	0.82	0.47	0.58	0.04
Resize	2DFMP	0.52	0.47	0.31	0.20	0.82	0.60	0.44	0.03
Resize	3DWavenet	0.47	0.48	0.29	0.20	0.81	0.53	0.47	0.00
Zeros	3D	0.61	0.49	0.32	0.21	0.83	0.60	0.47	0.05
Zeros	2D	0.60	0.49	0.32	0.21	0.83	0.60	0.47	0.05
Zeros	2DFMP	0.60	0.58	0.36	0.26	0.86	0.60	0.58	0.14
Zeros	3DWavenet	0.56	0.53	0.31	0.22	0.83	0.53	0.53	0.05

List of Figures

1.1	Coronary Artery Plaques [MH14]	2
1.2	Different plaque features and their shape in CCTA [Puc14]	4
1.3	Straightened CPR view of ill defined plaque segments	6
1.4	Workflow of noninvasive FFR [Min15]	7
2.1	Coronary Arteries[Cli18]	13
2.2	Radiomics Pipeline[Kol17]	15
2.3	GLCM calculation[Kol17]	17
2.4	GLRLM calculation[Kol17]	18
2.5	Decision Boundary	22
2.6	Support Vector Machines[Noe16]	23
2.7	Decision Tree	25
2.8	Multi-layer Perceptron[Mai18]	27
2.9	Activation Function	29
2.10	Convolutional Neural Network	30
2.11	Elman Recurrent Cell [Mai18]	31
2.12	WaveNet structure [vdO16]	32
3.1	Examples of Excluded Samples	38
3.2	Workflow of the Deep Learning Classification	43
4.1	Best Results Graph	51
4.2	Results compared to related work	53
B.1	3D Convolutions with WaveNet	68
B.2	2D Convolutions with GRUs	69
B.3	3D Convolutions with GRUs	70
B.4	2D Convolutions with Fractional Max Pooling and GRUs	71

List of Tables

A.1	Co-occurrence of Revascularisation and other Targets	65
A.2	Co-occurrence of Stenosis Grade and other Targets	65
A.3	Co-occurrence of Stenosis Grade with Revascularisation against other Targets	65
A.4	Co-occurrence of Spotty Calcifications and other Targets	66
A.5	Co-occurrence of Low HU Attenuation and other Targets	66
C.1	Best Results	73
C.2	Stenosis Degree	74
C.3	Positive Remodeling	74
C.4	Low HU Attenuation	75
C.5	Spotty Calcification	75
C.6	Revascularisation Plaque Level	76
C.7	Revascularisation Plaque Level Edgecases	76
C.8	Revascularisation Branch Level	77
C.9	Revascularisation Patient Level	77
C.10	Selected Features 1	79
C.11	Selected Features 2	80
C.12	Stenosis Degree	81
C.13	Positive Remodeling	81
C.14	Low HU Attenuation	82
C.15	Spotty Calcification	82
C.16	Revascularisation Plaque Level	83
C.17	Revascularisation Plaque Level Edgecases	83

Bibliography

- [Ach04] Stephan Achenbach et al. Assessment of coronary remodeling in stenotic and non-stenotic coronary atherosclerotic lesions by multidetector spiral computed tomography. *Journal of the American College of Cardiology*, 43(5):842–847, 2004.
- [Aga90] Arthur Agatston et al. Quantification of coronary artery calcium using ultrafast computed tomography. *Journal of the American College of Cardiology*, 15(4):827–832, 1990.
- [Ama89] Moses Amadasun and Robert King. Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(5):1264–1274, 1989.
- [AZ15] Armin Arbab-Zadeh and Valentin Fuster. The myth of the vulnerable plaque: transitioning from a focus on individual lesions to atherosclerotic disease burden for coronary artery disease risk assessment. *Journal of the American College of Cardiology*, 65(8):846–855, 2015.
- [Bai18] Shaojie Bai, Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.
- [Bal14] Yoganand Balagurunathan et al. Test–retest reproducibility analysis of lung CT image features. *Journal of Digital Imaging*, 27(6):805–823, 2014.
- [Bou17] Sabri Boughorbel, Fethi Jarray, and Mohammed El-Anbari. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PloS one*, 12(6):e0177678, 2017.
- [Bre01] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [Bur01] Allen Burke et al. Healed plaque ruptures and sudden coronary death: evidence that subclinical rupture has a role in plaque progression. *Circulation*, 103(7):934–940, 2001.

- [Cal11] Patrick Calvert et al. Association between IVUS findings and adverse outcomes in patients with coronary artery disease: the VIVA (VH-IVUS in Vulnerable Atherosclerosis) Study. *JACC: Cardiovascular Imaging*, 4(8):894–901, 2011.
- [Che16] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016.
- [Cho14] Kyunghyun Cho et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint: 1406.1078*, 2014.
- [Cla11] Bimmer Claessen et al. Impact of lesion length and vessel size on clinical outcomes after percutaneous coronary intervention with everolimus-versus paclitaxel-eluting stents: pooled analysis from the SPIRIT (clinical evaluation of the XIENCE V everolimus eluting coronary stent system) and COMPARE (second-generation everolimus-eluting and paclitaxel-eluting stents in real-life practice) randomized trials. *JACC: Cardiovascular Interventions*, 4(11):1209–1215, 2011.
- [Cli18] Cleveland Clinic. Heart & blood vessels: Your coronary arteries. Website, November 2018. accessed 28.11.2018.
- [Coe14] Adriaan Coenen et al. Fractional flow reserve computed from noninvasive CT angiography data: diagnostic performance of an on-site clinician-operated computational fluid dynamics algorithm. *Radiology*, 274(3):674–683, 2014.
- [Cur16] Ricardo Cury et al. Coronary artery disease-reporting and data system (CAD-RADS): an expert consensus document of SCCT, ACR and NASCI: endorsed by the ACC. *JACC: Cardiovascular Imaging*, 9(9):1099–1113, 2016.
- [Dav12] Fergus Davnall et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights into Imaging*, 3(6):573–589, 2012.
- [DG16] Jakob De Geer et al. Software-based on-site estimation of fractional flow reserve using standard coronary CT angiography data. *Acta Radiologica*, 57(10):1186–1192, 2016.
- [Dud12] Richard Duda, Peter Hart, and David Stork. *Pattern Classification*. John Wiley & Sons, 2012.
- [Elm90] Jeffrey Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990.

- [Fer12] Maros Ferencik et al. A computed tomography-based coronary lesion score to predict acute coronary syndrome among patients with acute chest pain and significant coronary stenosis on coronary computed tomographic angiogram. *The American Journal of Cardiology*, 110(2):183–189, 2012.
- [Fre97] Yoav Freund and Robert Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.
- [Fri01] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer Series in Statistics New York, NY, USA:, 2001.
- [Fus92] Valentin Fuster, Lina Badimon, Juan Badimon, and James Chesebro. The pathogenesis of coronary artery disease and the acute coronary syndromes. *New England Journal of Medicine*, 326(5):310–318, 1992.
- [Gal74] Mary Galloway. Texture analysis using grey level run lengths. *NASA STI/Recon Technical Report N*, 75, 1974.
- [Gra14] Benjamin Graham. Fractional max-pooling. *arXiv preprint arXiv:1412.6071*, 2014.
- [Gro82] Stephen Grossberg. Contour enhancement, short term memory, and constancies in reverberating neural networks. In *Studies of Mind and Brain*, pages 332–378. Springer, 1982.
- [Guy03] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3:1157–1182, 2003.
- [Har73] Robert Haralick et al. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, 3(6):610–621, 1973.
- [He16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [Hoc97] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [Iof15] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint: 1502.03167*, 2015.

- [Ip90] John Ip, Valentin Fuster, Lina Badimon, Juan Badimon, Mark Taubman, and James Chesebro. Syndromes of accelerated atherosclerosis: role of vascular injury and smooth muscle cell proliferation. *Journal of the American College of Cardiology*, 15(7):1667–1687, 1990.
- [Itu16] Lucian Itu et al. A machine-learning approach for computation of fractional flow reserve from coronary computed tomography. *Journal of Applied Physiology*, 121(1):42–52, 2016.
- [Kir13] Hortense Kirişli et al. Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography. *Medical Image Analysis*, 17(8):859–876, 2013.
- [Kol17] Márton Kolossváry et al. Radiomic features are superior to conventional quantitative computed tomographic metrics to identify coronary plaques with napkin-ring sign clinical perspective. *Circulation: Cardiovascular Imaging*, 10(12):e006843, 2017.
- [Kol18] Márton Kolossváry, Miklós Kellermayer, Béla Merkely, and Pal Maurovich-Horvat. Cardiac computed tomography radiomics: a comprehensive review on radiomic techniques. *Journal of Thoracic Imaging*, 33(1):26–34, 2018.
- [Koo11] Bon-Kwon Koo et al. Diagnosis of ischemia-causing coronary stenoses by noninvasive fractional flow reserve computed from coronary computed tomographic angiograms: results from the prospective multicenter discover-flow (diagnosis of ischemia-causing stenoses obtained via noninvasive fractional flow reserve) study. *Journal of the American College of Cardiology*, 58(19):1989–1997, 2011.
- [Kri12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.
- [Kru16] Mariusz Kruk et al. Workstation-based calculation of CTA-based FFR for intermediate stenosis. *JACC: Cardiovascular Imaging*, 9(6):690–699, 2016.
- [Kub10] Takashi Kubo et al. The dynamic nature of coronary artery lesion morphology assessed by serial virtual histology intravascular ultrasound tissue characterization. *Journal of the American College of Cardiology*, 55(15):1590–1597, 2010.

- [Lam12] Philippe Lambin et al. Radiomics: extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, 48(4):441–446, 2012.
- [LeC15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [Lib02] Peter Libby and Masanori Aikawa. Stabilization of atherosclerotic plaques: New mechanisms and clinical targets. *Nature Medicine*, 8(11):1257–1263, 2002.
- [Low14] Richard Lowry. *Concepts and applications of inferential statistics*. Commonwealth Of Learning, 2014.
- [Lug14] Felix Lugauer, Yefeng Zheng, Joachim Hornegger, and Michael Kelm. Precise lumen segmentation in coronary computed tomography angiography. In *International MICCAI Workshop on Medical Computer Vision*, pages 137–147. Springer, 2014.
- [Maa13] Andrew Maas, Awni Hannun, and Andrew Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. ICML*, volume 30, 2013.
- [Mai18] Andreas Maier, Nishant Ravikumar, Tobias Würfl, Katharina Breininger, and Vincent Christlein. Lecture notes in deep learning. University Lecture, October 2018.
- [Men15] Shanthi Mendis, Stephen Davis, and Bo Norrving. Organizational update: the World Health Organization global status report on noncommunicable diseases 2014; one more landmark step in the combat against stroke and vascular disease. *Stroke*, 46(5):e121–e122, 2015.
- [MH10] Pál Maurovich-Horvat et al. The napkin-ring sign: CT signature of high-risk coronary plaques? *JACC: Cardiovascular Imaging*, 3(4):440–444, 2010.
- [MH14] Pál Maurovich-Horvat, Maros Ferencik, Szilard Voros, Béla Merkely, and Udo Hoffmann. Comprehensive plaque assessment by coronary CT angiography. *Nature Reviews Cardiology*, 11(7):390–402, 2014.
- [Min15] James Min et al. Noninvasive fractional flow reserve derived from coronary CT angiography. *JACC: Cardiovascular Imaging*, 8(10):1209–1222, 2015.
- [Mot07] Sadako Motoyama et al. Multislice computed tomographic characteristics of coronary lesions in acute coronary syndromes. *Journal of the American College of Cardiology*, 50(4):319–326, 2007.

- [Mot09] Sadako Motoyama et al. Computed tomographic angiography characteristics of atherosclerotic plaques subsequently resulting in acute coronary syndrome. *Journal of the American College of Cardiology*, 54(1):49–57, 2009.
- [Nag03] Morteza Naghavi. From vulnerable plaque to vulnerable patient: a call for new definitions and risk assessment strategies. Part II. *Circulation*, 108:1772–1778, 2003.
- [Nak13] Ryo Nakazato et al. Noninvasive fractional flow reserve derived from computed tomography angiography for coronary lesions of intermediate stenosis severity: results from the DeFACTO study. *Circulation: Cardiovascular Imaging*, 6(6):881–889, 2013.
- [Noe16] Elmar Noeth, Joachim Hornegger, Dieter Hahn, and Stefan Steidl. Lecture notes in pattern recognition. University Lecture, October 2016.
- [Nør14] Bjarne Nørgaard et al. Diagnostic performance of noninvasive fractional flow reserve derived from coronary computed tomography angiography in suspected coronary artery disease: the NXT trial (Analysis of Coronary Blood Flow Using CT Angiography: Next Steps). *Journal of the American College of Cardiology*, 63(12):1145–1155, 2014.
- [O'C15] James O'Connor et al. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. *Clinical Cancer Research*, 21(2):249–257, 2015.
- [Par14] Chintan Parmar et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PloS one*, 9(7):e102107, 2014.
- [Pij96] Nico Pijls et al. Measurement of fractional flow reserve to assess the functional severity of coronary-artery stenoses. *New England Journal of Medicine*, 334(26):1703–1708, 1996.
- [Puc14] Stefan Puchner et al. High-risk plaque detected on coronary CT angiography predicts acute coronary syndromes independent of significant stenosis in acute chest pain: results from the ROMICAT-II trial. *Journal of the American College of Cardiology*, 64(7):684–692, 2014.
- [Ras06] Margaret Leila Rasouli, David Shavelle, William French, Charles McKay, and Matthew Budoff. Assessment of coronary plaque morphology by contrast-enhanced computed tomographic angiography: comparison with intravascular ultrasound. *Coronary Artery Disease*, 17(4):359–364, 2006.

- [Red16] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.
- [Ren14] Matthias Renker et al. Comparison of diagnostic value of a novel noninvasive coronary computed tomography angiography method versus standard coronary angiography for assessing fractional flow reserve. *The American Journal of Cardiology*, 114(9):1303–1308, 2014.
- [Ric89] Peter Richardson. Influence of plaque configuration and stress distribution on fissuring of coronary atherosclerotic plaques. *The Lancet*, 334(8669):941–944, 1989.
- [Ros58] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386, 1958.
- [Sch06] Joanne Schuijf et al. Relationship between noninvasive coronary angiography with multi-slice computed tomography and myocardial perfusion imaging. *Journal of the American College of Cardiology*, 48(12):2508–2514, 2006.
- [SG14] Carlos Santos-Gallego, Belen Picatoste, and Juan José Badimón. Pathophysiology of acute coronary syndrome. *Current Atherosclerosis Reports*, 16(4):401ff, 2014.
- [Sil16] David Silver et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.
- [Sun83] Chengjun Sun and William Wee. Neighboring gray level dependence matrix for texture classification. *Computer Vision, Graphics, and Image Processing*, 23(3):341–352, 1983.
- [Tay13] Charles Taylor, Timothy Fonte, and James Min. Computational fluid dynamics applied to cardiac computed tomography for noninvasive quantification of fractional flow reserve: Scientific basis. *Journal of the American College of Cardiology*, 61(22):2233–2241, 2013.
- [Tes17] Christian Tesche et al. Coronary CT angiography–derived fractional flow reserve. *Radiology*, 285(1):17–33, 2017.
- [Thi13] Guillaume Thibault et al. Shape and texture indexes application to cell nuclei classification. *International Journal of Pattern Recognition and Artificial Intelligence*, 27(01):1357002, 2013.

- [vdO16] Aaron van den Oord et al. WaveNet: A Generative Model for Raw Audio. *arXiv preprint:1609.03499*, 2016.
- [Ver13] Mathijs Versteylen et al. Additive value of semiautomated quantification of coronary artery disease using cardiac computed tomographic angiography to predict future acute coronary syndrome. *Journal of the American College of Cardiology*, 61(22):2296–2305, 2013.
- [vG17] Joost van Griethuysen et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Research*, 77(21):e104–e107, 2017.
- [Wel16] Michael Wels, Felix Lades, Christian Hopfgartner, Chris Schwemmer, and Michael Sühling. Intuitive and accurate patient-specific coronary tree modeling from cardiac computed-tomography angiography. In *The 3rd MICCAI WS interactive Medical Image Computing Workshop*, pages 86–93, 2016.
- [Wil90] Robert Wilson, Keith Wyche, Betsy Christensen, Steven Zimmer, and David Laxson. Effects of adenosine on human coronary arterial circulation. *Circulation*, 82(5):1595–1606, 1990.
- [Wil14] Martin Willemink et al. Coronary artery calcification scoring with state-of-the-art CT scanners from different vendors has substantial effect on risk classification. *Radiology*, 273(3):695–702, 2014.
- [Yan16] Dong Hyun Yang et al. Diagnostic performance of on-site CT-derived fractional flow reserve versus CT perfusion. *European Heart Journal–Cardiovascular Imaging*, 18(4):432–440, 2016.
- [Zhe13] Yefeng Zheng, Huseyin Tek, and Gareth Funka-Lea. Robust and accurate coronary artery centerline extraction in CTA by combining model-driven and data-driven approaches. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 74–81, Nagoya, 2013. Springer.
- [Zre18] Majd Zreik et al. Automatic detection and characterization of coronary artery plaque and stenosis using a recurrent convolutional neural network in coronary CT angiography. *arXiv preprint:1804.04360*, 2018.