Prediction of Obstructive Coronary Artery Disease using Machine Learning Algorithms

Rutger J. Metselaar

UNIVERSITY OF TWENTE.

ISALA

UNIVERSITY OF TWENTE

MASTER THESIS

Prediction of Obstructive Coronary Artery Disease using Machine Learning Algorithms

Author:
R. J. METSELAAR

Graduation committee:

Prof. dr. ir. C.H. SLUMP
Dr. J.D. VAN DIJK
Dr. J.A. VAN DALEN
Drs. B.N.VENDEL, MD
Drs. B.J.C.C. HESSINK-SWEEP
Dr. E. GROOT JEBBINK

A thesis submitted in fulfillment of the requirements for the degree of Master of Science

November 21, 2020

this page intentionally left blank

Summary

Introduction Accurate risk stratification for patients with coronary artery disease(CAD) is essential for accurate treatment. The current diagnostic pathway comprises a number of medical examinations , including a computed tomography scan and positron emission tomography myocardial perfusion, which yield prognostic data that may be utilized for risk stratification purposes. The aim of this thesis was to develop a risk model for obstructive CAD with machine learning(ML) algorithms. This model may provide an individualized risk score based on a combination of clinical features and quantitative parameters derived from imaging.

Methods We retrospectively included 1007 patients with no prior cardiovascular history, who were referred for rest and regadenoson-induced stress Rubidium-82 positron emission tomograpy (PET)/computated tomography (CT). Presence of obstructive CAD was defined as a composite of a significant fractional flow rate measurement during invasive coronary angiography, percutaneous coronary intervention or a coronary artery bypass graft procedure, and was acquired via follow-up. Furthermore, each patient was characterized by a broad array of features, including cardiovascular risk factors (cigarette smoking, hypertension, hypercholesterolemia, diabetes, positive family history of CAD), prior medical history; current medication usage age; gender; body mass index (BMI); creatinine serum values; coronary artery calcification (CAC) score and PET/CT derived myocardial blood flows. Additionally, the visual interpretation by a team of two clinicians of the PET/CT scan was obtained. Two sets of input parameter were used to train the models. First, the entire set of features except the visual interpretation. Secondly, the entire set of features, including the visual interpretation. Four different ML algorithms were used, so in total, 8 different models were optimized. These models were developed using a subset of 805 cases of the dataset to identify obstructive CAD by using 5-fold cross validation in combination with a grid search, whilst their performance was measured using the F1-score. The optimized algorithms were validated on 202 cases of the dataset, never previously seen by the models. The performance on these unseen examples was compared with the current diagnostic performance by clinicians, as measured by the visual interpretation of the scan.

Results The best performing algorithm to predict obstructive CAD was XGBoost, an ensemble of gradient boosted decision trees. On the unseen dataset this algorithm reached an area under the curve of 0.93 while obtaining a sensitivity of 64% (95% CI: 41-83) and a specificity of 96% (95% CI: 91-98). The sensitivity by the clinicians on this same dataset was 77% (95% CI 55-93) and the specificity was 92% (95% CI (87-96). The low prevalence of obstructive events in evaluation dataset (11%) resulted in wide confidence intervals, making it so that no significant differences were found. Furthermore, we were able to make a ranking via the XGBoost model of important predictors for obstructive CAD. Summarized, CAC-scores and quantitative

PET derived features were the most important predictors. Classical risk factors and medication however, could not be used in the current setup to distinguish obstructive CAD from non-obstructive CAD. We also conclude that the visual interpretation by the clinician added incremental prognostic information to the model.

Conclusion We used a set of clinical and quantitative features to develop a ML model. This model is able to provide individualized risk stratification by predicting the possibility of an obstructive cardiovascular event. Although validation with a larger dataset could result in a more well defined performance range, this model still shows potential to be implemented in the diagnostic workflow by providing a computer aided second opinion to the clinicians.

Acknowledgements

'Asking for help is not giving up, it is refusing to give up'. Deze wijsheid is blijven hangen nadat ik hem onlangs in een boek heb gelezen. Laat ik maar hopen dat dit daadwerkelijk zo is, want zonder de hulp van een groot aantal mensen had ik mijn onderzoek nooit kunnen doen. In het bijzonder wil ik de volgende personen bedanken:

Allereerst, veel dank gaat uit naar mijn begeleiding in het Isala: Joris, Jorn en Brian. Ik kon altijd bij jullie terecht met mijn problemen, ongeacht van welke aard. Jullie vakkennis, geduld, enthousiasme voor het onderzoek en vriendelijkheid heb ik enorm kunnen waarderen. Ik kan me werkelijk geen betere begeleiding voorstellen. Bedankt voor alles.

Kees, mede dankzij jou is het gelukt om het niveau van mijn thesis naar een hoger resultaat te brengen. Ik waardeer het dat je tot in de kleinste details (en met humor) scherpe feedback weet te geven. Erik, erg fijn dat je als buitenlid in mijn afstudeercommissie wilde aanschuiven. Het treft dat jouw deelname tegelijkertijd wat Deventer's tegenwicht biedt.

Bregje, ik wil je graag bedanken voor de inzichten die je me gedurende al mijn stages hebt kunnen geven. Ik heb de afgelopen tijd veel over mezelf geleerd, en ben mezelf ook zeker tegengekomen. Dankzij jouw adequate bijsturing tijdens intervisies ben ik ondanks deze aparte corona tijd op het juiste spoor gebleven.

Maar een intervisie is niet compleet zonder mede lotgenoten. Thijs, Tess en Friso veel dank voor jullie aanmoediging en thuiswerk tips tijdens deze vreemde coronatijd. Ik wens jullie het beste!

Verder wil ik Ludo Cornelissen van het UMCG bedanken voor zijn inzichten en assistentie met betrekking tot Machine Learning, en Henkjan Huisman van DIAG voor het inzetten van zijn netwerk om mij op de goede weg te helpen.

Sabine, Mark en Jorik, bedankt voor alle gesprekken en gezelligheid in het Isala. We hebben samen met elkaar gelachen, gespard maar ook hard kunnen werken. Zonder jullie was mijn tijd in het Isala een stuk minder plezierig geweest!

En niet het minst wil ik mijn ouders bedanken. Anja en Dirk, jullie eeuwige steun waardeer ik enorm. Ook bij tegenslagen hebben jullie mij iedere keer weten te motiveren om er wat van te maken. Dankzij jullie toewijding heb ik van dit laatste jaar een succes kunnen maken.

Veel plezier met het lezen van deze thesis,

Hartelijke groet,

Rutger

Deventer, 20 November 2020

Contents

List of Abbreviations

CAD	**C**oronary **A**rtery **D**isease
ML	**M**achine **L**earning
PET	**P**ositron **E**mission **T**omography
CT	**C**omputated **T**omography
BMI	**B**ody **M**ass **I**ndex
CAC	**C**oronary **A**rtery **C**alcification
AI	**A**rtificial **I**ntelligence
CD	**C**ardiovascular **D**isease
AP	**A**ngina **P**ectoris
MI	**M**yocardial **I**nfarction
LM	**L**eft **M**ain coronary artery
LAD	**L**eft **A**nterior **D**escending coronary artery
LCX	**L**eft **C**ircumflex coronary artery
RCA	**R**ight **C**oronary **A**rtery
ESC	**E**uropean **S**ociety of **C**ardiology
PTP	**P**re-test **P**robability
CTA	**C**omputated **T**omography **A**ngiography
HU	**H**ounsfield **U**nit
NPV	**N**egative **P**redictive **V**alue
ICA	**I**nvasive **C**oronary **A**ngiography
MPI	**M**yocardial **P**erfusion **I**maging
SPECT	**S**ingle **P**roton **E**mission **C**omputated **T**omography
MBF	**M**yocardial **B**lood **F**low
MFR	**M**yocardial **F**low **R**eserve
FFR	**F**ractional **F**low **R**eserve
PCI	**P**ercutaneous **C**oronary **I**ntervention
CABG	**C**oronary **A**rtery **B**ypass **G**rafting
LR	**L**ogistic **R**egression
LASSO	**L**east **A**bsolute **S**hrinkage and **S**election **O**perator
SVM	**S**upport **V**ector **M**achine
XGB	e**X**treme **G**radient **B**oosting
MSE	**M**ean **S**quared **E**rror
ROC	**R**eceiver **O**perator **C**haracteristic
AUC	**A**rea **U**nder the **C**urve
DT	**D**ecision **T**ree
ROI	**R**egion **O**f **I**nterest
TACs	**T**ime **A**ctivity **C**urves
EF	**E**jection **F**raction
SSS	**S**ummed **S**tress **S**cores
SDS	**S**ummed **D**ifference **S**cores
MPE	**M**yocardial **P**erfusion **E**ntropy
MACE	**M**ajor **A**dversarial **C**ardiac **E**vent
CNN	**C**onvolutional **N**eural **N**etwork

Chapter 1

General Introduction

The study of Technical Medicine is conceptualized to introduce and implement technological advances in clinical practice. During the last decades one particular technological field took on an increase of interest: Artificial Intelligence (AI) and its' subfield machine learning (ML). Advances not only affected the theoretical aspects. The practical generation and deployment of ML models has become more accessible too. Several open source ML frameworks, exponential advances in hardware technology and the digitization of companies and data are all contributing factors.

In the healthcare industry too, the adaptation of ML exploded. In hospitals this is visible by the rapid increase of ML related applications, especially in radiology and intensive care departments. The potential benefit of these applications have led Isala hospital to include AI into their strategy for the coming years. In Isala, which is one of the largest general hospitals in The Netherlands, the most important resources required to develop, refine and validate ML applications are available: High volume and high quality patient data. Combined with Isala's neverlasting aspiration to improve disease diagnosis and treatment for patients, the potential benefit of employing the power of ML becomes clear.

The thesis of my final internship will cover this: The development of a clinically relevant tool with ML. In my case this is a tool that should be able to assist cardiologists in Isala in cases where the clinical decision is not straightforward. More specifically, I have developed a predictive model for patients with coronary artery disease (CAD) that can assist clinicians by providing a risk assessment of a patient. This risk assessment should be based on prognostic factors. Some of these factors are already well understood. However, other features may not always seem obvious nor intuitive to a clinician, but are expected to be influential nonetheless.

1.1 Outline of this thesis

The first chapter describes the clinical relevance of CAD, including the relevant clinical background. Also, the problem statement and the aim of this thesis will be described in greater detail. In the second chapter an introduction on ML is given in the context of the problem statement. Four different ML algorithms will be highlighted. In the third chapter we focus on the development and validation of ML techniques for the prediction of obstructive CAD. The fifth and final chapter discusses future perspectives, clinical implementations and recommendations for future research.

Chapter 2

Clinical Background

2.1 Introduction on Obstructive Coronary Artery Disease

Cardiovascular disease (CD) is an umbrella term for all disease affecting the circulatory system or the heart[1]. CD is the most common cause of death, representing almost 34% of all global deaths in 2015. In developed countries the mortality due to CD is lower compared to less developed countries due to proper clinical management and health campaigns, but the burden on healthcare caused by CD remains significant. Approximately 4.4% of the Dutch population is known to suffer from some form of CD in 2019, and most of these people receive some form of therapy[2]. Coronary artery disease (CAD) is responsible for nearly half of the CD related deaths making it the single largest contributor to mortality[3].

CAD involves progressive narrowing (stenosis) and consequently obstruction of the coronary arteries, resulting in a reduction of oxygenation of the myocardium. The progressive process may eventually lead to an ischemia or infarction[1]. The disease that involves the narrowing of arteries by build-up of plaques is atherosclerosis, which is often the culprit of CAD. So-called atheromatous plaques consist of fat, cholesterol and calcium and other substances found in blood. Progression of the plaques is twofold. Over time plaques grow in size and result in a reduction of the lumen of an artery. Additionally gradual calcification of plaques occurs due to a process that resembles bone formation[4].

Patients with atherosclerosis can remain asymptomatic for decades due to the progressive nature of the disease. Although arteries are able to compensate for the narrowing by growing larger in diameter, a process known as arterial remodeling, this remodeling can only compensate to an extent[5]. Symptoms start to arise when stenosis cause a critical reduction of blood flow to organs. For patients with CAD, the most typical symptom is stable angina pectoris(AP), a type of chest pain that is the result of an imbalance between myocardial oxygen supply and demand during stress. The myocardium then becomes ischemic. Even more drastic is a complete obstruction of an artery, also known as a myocardial infarction (MI). MI typically follows ulceration of atheromatous plaques causing a cascade of events involving local blood clotting.

2.2 Anatomy

Three major arteries within the coronary circulation that supply the myocardium of nutrients and oxygenation can be defined. The first two arise from the left main (LM) coronary artery, which originates from just above the aortic valves: the left anterior

descending (LAD) artery and the left circumflex (LCX) artery. The third coronary artery is the right coronary artery (RCA) and originates directly from the aorta[6].

Most of the coronary blood flow is dedicated to supply the left ventricle, the part of the heart that contracts as to pump blood to the rest of the body. Hence, myocardial ischemia is the most severe when blood supply to the left ventricle becomes impaired. The anatomy and the perfusion of the left ventricle can be represented in a standardized method with the 17-segment model of the American Heart Association as shown in Figure 2.1[7]. Although there is some natural variability in coronary anatomy within humans, this model standardizes vascular territories for most people[6, 7].

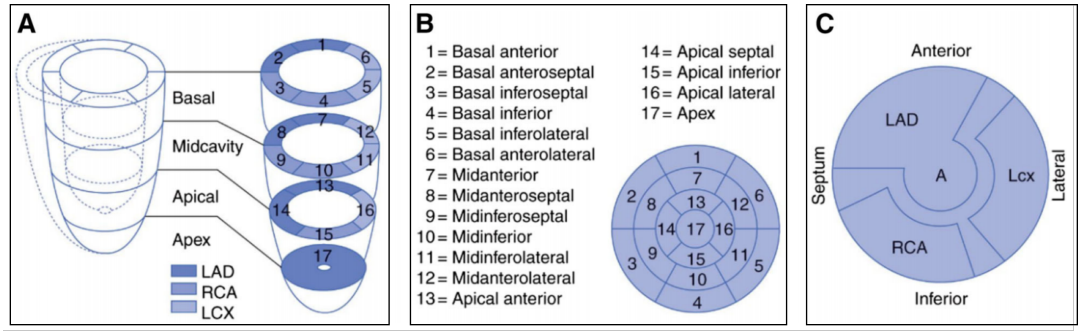

FIGURE 2.1: Myocardial segmentation and standard nomenclature from the American Heart Association. (A) The standard segmentation model divides the myocardium into three major short-axis slices: apical, mid-cavity and basal. (B) The segments can be flattened into a bulls-eye. (C) Approximate perfusion regions for the three major coronary arteries can be projected onto this bulls-eye. Adapted from Dilsizian et al.[8].

2.3 Diagnosis of Obstructive Coronary Artery Disease

Due to the progressive nature of CAD and its long-term consequences, early detection is crucial for patients. Symptoms such as AP are often the reason patients present themselves to the clinician. The latest European Society of Cardiology (ESC) guidelines propose a six-step approach for the management of patients with angina pectoris and suspected CAD, as shown in Figure 2.2. In the first step, symptoms and signs are assessed to confirm that a patient is not suffering from unstable AP or acute myocardial infarction and in need of immediate revascularization. In the second step, the general condition and quality of life of a patient is evaluated. Comorbidities that may affect decision making are considered. Subsequently patients are followed up with basic testing such as ECG, biochemistry, echocardiography and X-ray. The cardiac function, characterized by wall movements and ejection fraction, of the left ventricle is assessed. Next, the clinical likelihood or the pre-test probability (PTP) of obstructive CAD is estimated using predictive risk models. Based on the PTP a type of additional diagnostic testing is recommended. Once the diagnosis of CAD is confirmed the patient's event risk is considered and this risk determines therapy options[9].

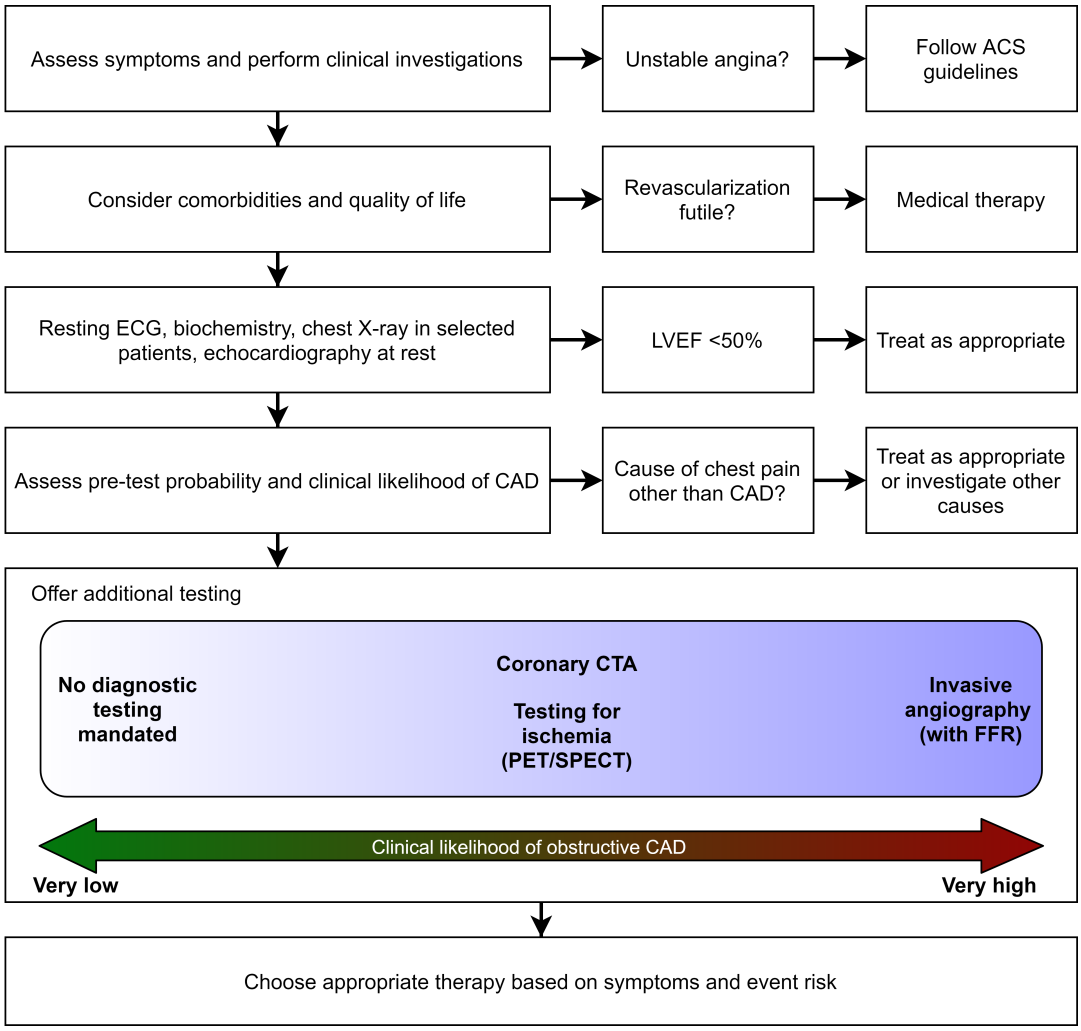

FIGURE 2.2: Schematic of the diagnostic path for patients suspected of CAD, which comprises a six-step approach. In the first step, symptoms are assessed. In case of unstable AP, the acute coronary syndrome(ACS) guidelines should be followed. With the second, comorbidities should be considered. If there are no revascularization options, appropriate medical therapy is recommended. During the third step a patient is to undergo additional examinations such as electrocardiogram(ECG), chest X-ray, biochemistry testing. Biochemistry testing includes laboratory tests focused on screening for atherosclerosis, tests focused on ischemea detection, tests for other cardiomyopathies, but also more general blood tests. If during these examinations the left ventricle ejection fraction(LVEF) is not found to less than 50%, the PTP is determined. Additional testing follows, depending on the clinical likelihood of obstructive CAD. For low risk patients, no diagnostic testing is mandated. For medium risk patients coronary computated tomography angiography(CTA) or myocardial perfusion imaging(MPI) is recommended. For high risk patients invasive angiography, with optionally a fractional flow reserve(FFR) measurement, is recommended. The final therapy should be chosen appropriately depending on symptoms and event risk.

2.4 Diagnostic pathway of obstructive coronary artery disease in Isala

More specific to my internship here in Isala, I will describe to some extent the diagnostic pathway for the detection of obstructive CAD, which deviates from the ESC guidelines.

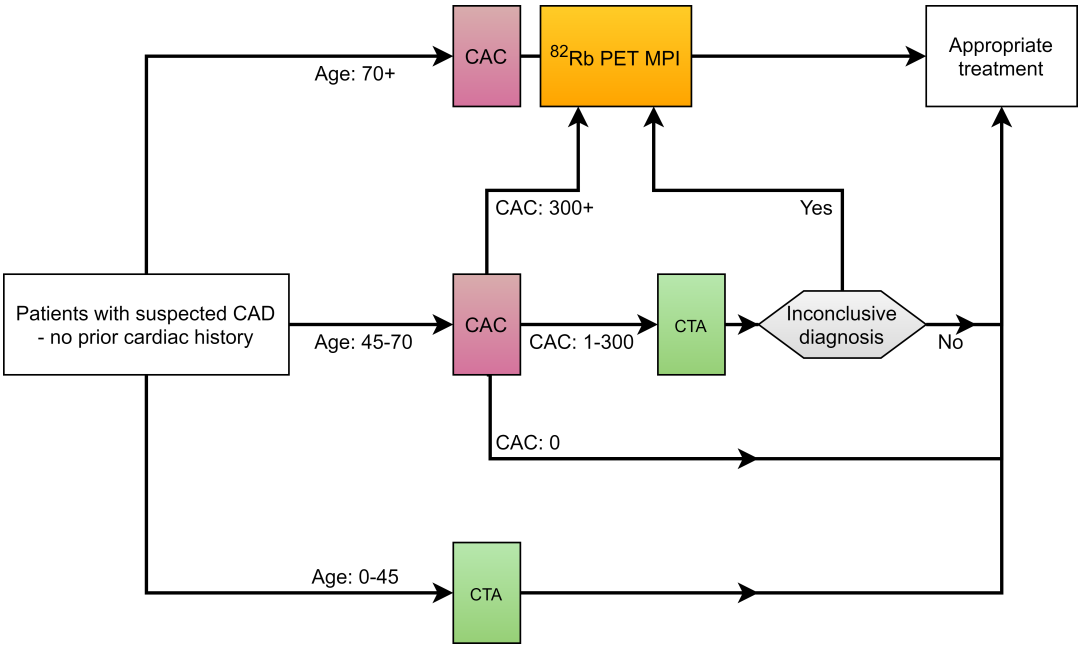

FIGURE 2.3: Workflow of the diagnostic pathway for patients with suspected CAD in Isala Hospital. The initial examination depends on the age of a patient. Patients under the age of 45 are referred for computated tomograpy angiography (CTA) because the likelihood of presence of calcification is very low, therefore making CTA the preferred choice. For patients over the age of 45 a CAC scan is made. For patients over 70 this is usually followed up with a 82Rb PET scan. For patients between 45 and 70 years old, the outcome of the coronary artery calcification (CAC) scan determines if additional examination is required, and which modality is preferred.

A simplified workflow of the diagnostic workflow for CAD in Isala is shown in Figure 2.3. There are multiple diagnostic pathways possible for patients, depending for example on when and where physical complaints are first diagnosed.

2.4.1 CAC-Score

Risk stratification is a cornerstone in the current diagnostic workup for patients with suspected CAD. A multitude of diagnostic modalities exist that can be used to investigate symptoms, and guidelines based on risk stratification provide guidance to clinicians towards appropriate diagnostic modalities[10]. In Isala patients with no relevant past medical history undergo coronary artery calcification (CAC) scoring. The CAC-score is derived from a unenhanced low-dose computed tomography (CT) scan and is a quantitative measure that is indicative of the severity of atherosclerosis in the coronary arteries[11]. The CAC-score is the result of a weighted sum of the lesions that contains calcification as visualized on the CT scan and is expressed in dimensionless Agatston Units[12]. The weight of the lesion depends on the density factor. The density factor follows directly from the highest plaque attenuation which is expressed in Hounsfield unit (HU) values. The relationship between density factor and HU is shown in Table 2.1.

General clinical interpretation of the value of the CAC-score is shown in Table 2.2[11]. CAC-scoring is especially useful for risk stratification because of the excellent negative predictive value (NPV) to detect CAD. The NPV is reported to be

TABLE 2.1: Density factor lookup table

Hounsfield Unit range (HU)	Density factor
130-199	1
200-299	2
300-399	3
400+	4

between 93% and 99% and a zero-score is associated a very low prevalence of obstructive CAD. Usually in these cases no clinical intervention is required[13]. High CAC-scores are correlated with an increased risk of obstructive CAD[14]. In order to distinguish between obstructive and non-obstructive CAD the consensus recommends additional diagnostic examinations in patients with a CAC-score between 1 and 400 AU[11]. Additionally the CAC-score is a useful measure that assists in the selection for additional diagnostics[9, 11].

TABLE 2.2: Clinical interpretation of the CAC-score

Degree of Coronary Artery Calcification	Absolute CAC-score	CAC-score adjusted for gender, age and ethnicity – percentile	Clinical Interpretation
Absent	0	0	Very low risk of future coronary events
Discrete	1-100	75	Low risk of future coronary events; low probability of myocardial ischemia
Moderate	101-400	76-90	Increased risk of future coronary events consider reclassifying the individual as high risk
Accentuated	>400	>90	Increased probability of myocardial ischemia

The National Institute for Health and Care Excellence guidelines propose that a CAC-score of more than 400 should be followed up with additional examinations like invasive coronary angiography (ICA)[15]. A less invasive approach would be to use functional imaging, such as myocardial perfusion imaging (MPI) with 82Rb-positron emission tomography (PET).

2.4.2 CT Angiography

For patients with a CAC-score between 1 and 300, CT Angiography (CTA) is the preferred option for additional examination of the coronaries. In case of higher CAC-scores the quality of the scan cannot be guaranteed due to significant blooming artifacts[16]. An Iodine based contrast agent is intravenously administered prior to the scan and results in enhanced contrast between the coronaries and surrounding tissue. The CTA scan results in a 3D volume that allows for anatomic interpretation as well as evaluation of the trajectory of the lumen of the arteries and of the condition of the vessel walls. With additional post processing each of the main vessels can be investigated more thoroughly. CTA is a relatively inexpensive and non-invasive modality that is highly accurate for the detection of CAD, with a reported pooled sensitivity and specificity of 66% and 89% respectively[17, 18].

2.4.3 Myocardial Perfusion Imaging

If CTA is not an option, the quantification and visualization of the perfusion pattern of the heart can provide insights in the origin and severity of the CAD related symptoms[19]. A well established method for quantification of myocardial perfusion is MPI with PET or single proton emission computated tomography(SPECT). Both modalities enable accurate non-invasive detection of CAD related perfusion defects[20]. This includes stenosis related defects but also irreversible defects including previous infarcts or scarred tissue[21, 22]. In a comparative study by Danad et al. PET was found to exhibit the highest accuracy for diagnosis of myocardial ischemia[23]. Ghotbi et al. raised a number of different advantages of PET over SPECT, including improved image quality, less radiation dose to patient and staff, rapid scan times and higher diagnostic accuracy[24].

In Isala, PET is the preferred modality and Rubidium-82 (^{82}Rb) is used as radiotracer. After intravenous injection of a saline solution containing ^{82}Rb, the radiotracer is distributed throughout the body. The interactions of ^{82}Rb within the human body are similar to those of potassium ions and the radiotracer is being actively transported within the myocardium by the sodium-potassium pumps. The uptake of 82Rb by the myocardium is related to perfusion and therefore viable myocardial cells can be distinguished from infarcted or necrotic tissue by differences in regional tracer uptake[25].

Cardiac complaints often become more severe during exercise since the myocardium requires more oxygen (and other nutrients). With MPI, exercise or 'stress' can be simulated using pharmaceuticals such as Regadenson or Adenosine[26]. These pharmaceuticals activate receptors for vasodilatation and this results in an increase in myocardial blood flow (MBF). In Isala Regadenoson is preferred since it selectively activates A2A receptors, as opposed to other pharmaceuticals that non-selectively activate the A1, A2B and A3 receptor subtypes. Nonselective activation results in negative chronotropic, dromotropic, inotropic, and anti-beta-adrenergic effects[27].

During MPI the cardiac perfusion is imaged during two phases: The rest phase and the pharmacologically induced stress phase. Defects in perfusion that are visible in both phases indicate previously damaged myocardium or scarred tissue. Perfusion defects that are more prominent during stress are indicative of ischemia.

Multiple methods have been proposed to quantify the myocardial blood flow from MPI, ranging from retention models, two compartment models and a one-compartment model, the latter one being used in Isala[25, 28–31]. This one-compartment model uses ^{82}Rb kinetics and a nonlinear extraction function to obtain estimates of MBF in normal myocardium[31]. The ratio between rest MBF and stress MBF is defined as the myocardial flow reserve(MFR). In combination with the 17-segment cardiac model of the left ventricle MBF and MFR can be characterized globally, per vessel and even more regionally per segment.

Since abnormal myocardial perfusion is a predictor for future cardiovascular events, normal ranges for MBF and MFR have been established[19, 32]. These values were calculated as the weighted mean from eight different studies on healthy subjects (mean age 28.6) with a total sample population of 382. For the one-compartment model, the average flow values of rest MBF are 0.74 mL/minute/g (range 0.69–1.15)

and the average flow values for stress MBF are 2.86 mL/minute/g (range 2.5-3.82). The weighted mean MFR was 4.07 (range 3.88-4.47)[32]. It can be seen that there is significant intervariability. But in general, low flow values in stress (<1.8) and low MFR values (<1.5) correlate with a negative prognosis for CAD. Naya et al. and Murthy et al. concluded that normal MFR had a high NPV for excluding high-risk CAD, and an abnormal MFR shows an increased probability of significant obstructive CAD[32, 33].

Overall, Regadenoson PET MPI has a sensitivity of 92% (95% CI 83%-97%) and a specificity of 77% (95% CI 66%-86%) for significant obstructive CAD[34]. This performance, combined with the non-invasive aspect of PET have both contributed to the technique becoming an essential element in the diagnostic workflow for CAD in Isala.

2.4.4 Invasive Coronary Angiography

ICA is the overall reference standard for the detection of a significant stenosis[23]. ICA is a minimally invasive procedure that allows visualization of the coronary arteries. This involves catheterization and injection of a contrast agent in the coronary arteries and subsequent dynamic imaging using X-rays. The same catheter can be used for pressure measurements in case of intermediate lesions. With pressure measurements the so-called fractional flow reserve (FFR) can be calculated over a stenosis to determine if the obstruction is functionally significant. This measure is the ratio of pressures measured over the stenosis, and is often utilized during ICA to determine the need of intervention and also to evaluate the result after stenting of ballooning stenosis.

The absolute diagnosis of significant obstructive CAD can be made with ICA. However, for diagnostic purposes, ICA is often the last resort not preferred as first option since the procedure is minimally invasive and therefore still accompanied with risks. On the other hand, if the likelihood of significant CAD is high, ICA comes with the benefit of immediate treatment possibilities by stenting.

2.4.5 Management of Patients with Coronary Artery Disease

In general, the management of CAD aims to reduce the symptoms and improve the prognosis through lifestyle changes accompanied with medical therapy and sometimes revascularization[9].

This requires a personalized approach and involves risk stratification. For patients classified as low risk of developing a cardiac event life style changes and sometimes medication suffice as treatment. Lifestyle changes aim to reduce risk factors. This involves smoking cessation, a healthy diet, physical activity and maintaining a healthy weight. Medication is used to reduce risk factors to prevent disease progression, but mostly for event prevention[9]. For patients diagnosed with CAD and who are classified as high risk, revascularization is suggested on top of medical therapy. Revascularization can be performed minimally invasive via percutaneous coronary intervention (PCI) or surgically via coronary artery bypass grafting (CABG).

2.5 Potential Improvements in the Diagnostic Workflow

The CAC-score, the CTA scan and the PET-MPI scan may all provide conclusive diagnosis for patients with suspected CAD. If these examinations indicate the need for intervention or are inconclusive, patients are usually referred for ICA. For many patients, however, the diagnosis is not definite after these examinations.

Out of 177 patients who were referred for ICA after PET MPI between May 2017 and January 2019, 111 were diagnosed with significant obstructive CAD. This means that 37% of patients underwent a procedure that could potentially have been omitted. Obviously this is insurmountable in order to reduce the number patients that have significant lesions, and it can be logically explained because the consequences of missing a potential case of high risk obstructive CAD (and the associated prognosis) outweigh the risks of ICA.

The process of risk stratification can be complicated in some cases. Figure 2.4 shows some of the most important variables that a clinician can use to evaluate the risk for a patient. It often remains difficult to interpret all these variables and put them in the correct context of the patient to estimate its risk-profile for cardiac events in the near future. First of all, there are simply too many variables. Secondly, some of these variables can be contradictive. It is known that most of these variables are interrelated via complicated nonlinear relationships, making them difficult to interpret for a clinician.

It has been suggested that advanced risk prediction models can assist the clinician in these difficult cases[9]. Prior work that incorporates these variables in risk prediction models is promising and showed the potential of integrating imaging derived features with clinical data in risk prediction models for improved risk stratification[35–37]. Machine learning (ML) algorithms were used to develop these models. This had led to the idea that for Isala, a similar risk prediction model can be developed with ML.

2.6 Aim of this Thesis

This thesis focuses on the use of ML for risk stratification of patients with obstructive CAD. One aspect of this project comprises the development of ML models, including a comparison to find the best performing approach. Another aspect is dedicated to the clinical evaluation and potential implementation. Since PET MPI is the standard for cardiac blood flow quantification, the aim is to combine PET MPI derived features together with a range of clinical features and use ML to obtain individual risk stratification for obstructive CAD.

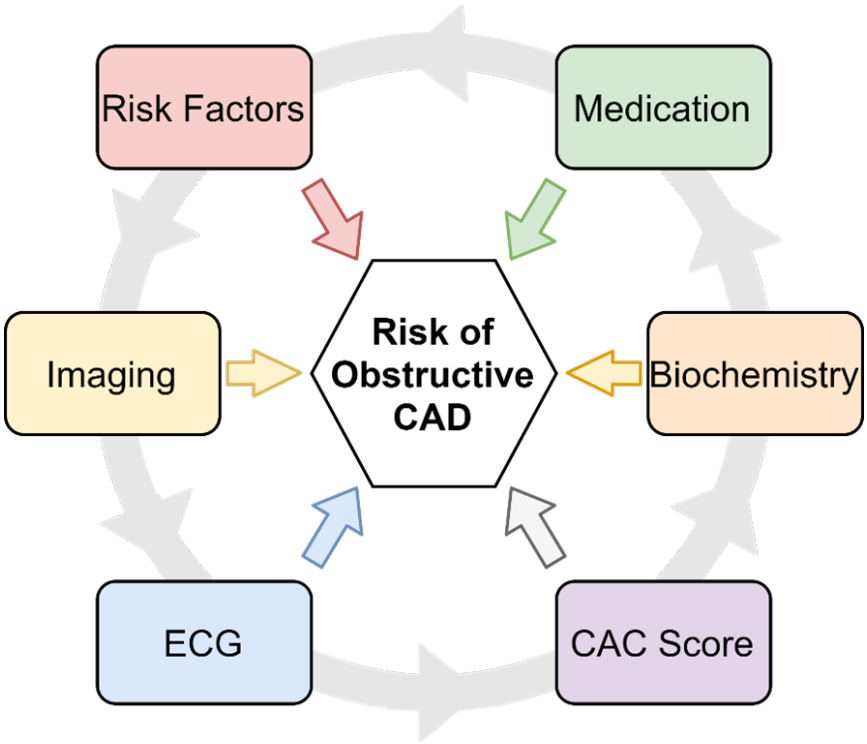

FIGURE 2.4: Variables that can be interpreted by a clinician to evaluate the risk of having an cardiovascular event for a patient. This includes classical risk factors for CAD, medication, imaging, including CAC-scores, CTA, and PET/SPECT MPI, biochemistry characteristics and ECG readings.Not only do these variables affect the risk of obstructive CAD, but they may also alter the interpretability of other risk elements. Hence, interpretation of these elements is not always straight forward.

Chapter 3

Technical Background

3.1 Machine Learning

ML is a discipline that falls under the umbrella term Artificial Intelligence(AI). It involves the study of computer algorithms that are trained to execute some task, and automatically improve in their ability to do so through experience. These algorithms can be viewed as mathematical functions that attempt to map an input vector $\mathbf{X}$ to a desired output vector $\mathbf{Y}$, and experience can be seen as the process of iteratively trying to improve on the performance of a task by modifying and fine-tuning parameters within the mathematical function $f(\mathbf{X})$. These tasks can be understood as optimization problems where the so-called loss, which is a value that represents the error, is minimized.

Four main categories of learning problems can be distinguished. The method of learning depends on the available data and the type of problem. The first category is supervised learning, and describes learning from data points where both input $\mathbf{X}$ and output $\mathbf{Y}$ are available. This entails learning from example pairs, or a 'labeled' dataset. The second category is unsupervised learning, where only input data is available. Unsupervised learning refers to the analysis of a dataset without knowing a priori what should be learned. It involves finding patterns or features, or clustering of data e.g. finding patterns in gene expressions. Finally there is reinforcement learning. In reinforcement learning, there is no direct access to the correct output, but it is possible to get a measure of the quality of output $\mathbf{Y}$ following input $\mathbf{X}$. For example, learning a car to drive within a setting that contains rules, and the distance that the car is able to drive is a measurement of quality. A fourth category that can be defined is transfer learning. In transfer learning, a machine exploits knowledge that it has learned from a previous task, and applies this information to another problem. For example, a model that has learned to recognize digits can be used to develop a model that recognizes characters[38].

3.1.1 General Principles

The optimization process of function $f(\mathbf{X})$ can be divided in three phases. Each phase requires independent examples of the available dataset, and therefore the dataset is split into parts. Each sub dataset is then used in one of the phases. The optimization process starts in the training phase and is followed by the validation and testing phase. Performance metrics are used to evaluate the performance of a model during these phases. To gain deeper understanding of ML, some of the principles explained in the following section will be illustrated with an example dataset

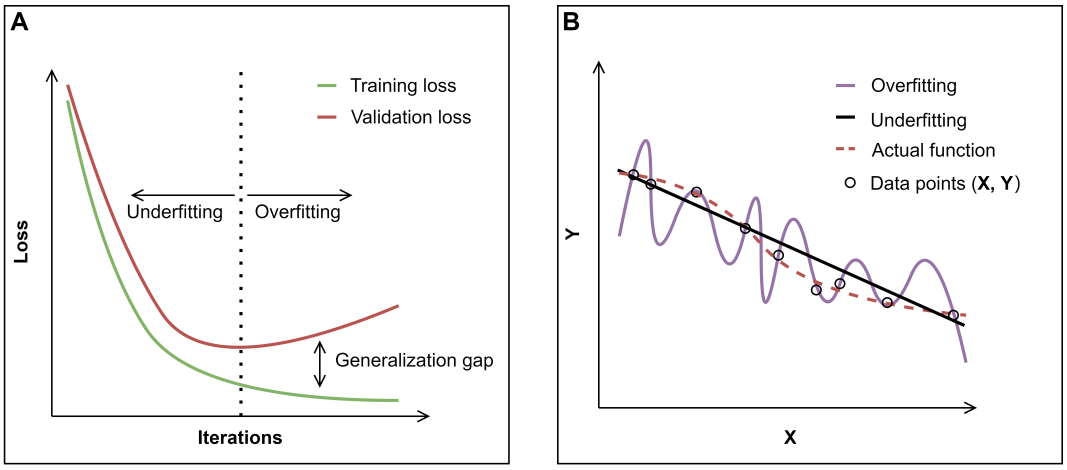

FIGURE 3.1: A) This graph shows the relationships between iterations, underfitting and overfitting and the corresponding generalization gap. B) Actual data points (X,Y) are plotted and several functions are implemented to fit the data points. In the case of overfitting, the accompanied loss is very low, however the function tends to misrepresent the actual function. On the other hand, in the case of underfitting, the variance is low and the accompanied bias is high, as well as the loss. The optimal model sits between underfitting and overfitting and should approximate the actual function.

that contains real-world data.

3.1.2 Performance Metric

The performance metric describes how well the function $f(\mathbf{X})$ performs a task T, thus giving an indication how well a model performs. The performance metric is essential for ML because this metric is usually incorporated in the feedback system of an algorithm, that is, each iteration (e.g. the processing of the entire dataset) parameters within the function $f(\mathbf{X})$ will be modified to improve the performance metric or to reduce the loss until the performance reaches an optimum. This enables the algorithm to 'learn' with experience. This process is visualized in Figure 3.1.

3.1.3 Development Phases while Training a Model

Phase 1: Training

During this phase the function F is 'trained' to map X to Y. This process is shown in Figure 3.1. However, you want this performance not to be exclusive to the dataset $\mathbf{X}$. The aim is to reach a similar performance on new or unseen data examples. That is why only a subset of the original dataset $(\mathbf{X})$ is used to optimize $f(\mathbf{X})$, while the remainder of the dataset is used to evaluate the actual performance of $f(\mathbf{X})$ on unseen data. It is common practice to use the majority (60-80%) of the dataset for this training phase. However, the proportion of the training, validation and testing dataset are largely dependent on the size of the dataset. The partitioning should be chosen in a manner to The risk is that $f(\mathbf{X})$ corresponds too closely or exactly to the dataset $(\mathbf{X}, \mathbf{Y})$ and fails to fit new data reliably. This is overfitting: $f(\mathbf{X})$ exploits apparent relationships that do not hold outside of the original dataset $(\mathbf{X}, \mathbf{Y})$.

This principle results in a subtle balance between underfitting the data and overfitting the data. Generalization is the term used to describe the ability of a model to interpret new or unseen examples, and this ability of $f(\mathbf{X})$ to generalize is evaluated later during the testing phase. Sometimes specific regularization terms are added to $f(\mathbf{X})$ to prevent overfitting. These regularization terms attempt to reduce overfitting, and are one of the tunable hyperparameter of $f(\mathbf{X})$. Hyperparameters characterize $f(\mathbf{X})$ and can be modified to control the learning process. Hyperparameters vary per algorithm and optimal hyperparameters vary per dataset. The subsequent validation phase is used to identify optimal hyperparameters for a task and to evaluate (and prevent) overfitting. Regularization generally refers to techniques that aim at reducing overfitting during training and are an essential component of ML.

Phase 2: Validation

The aim of the validation phase is to create a model that is specific but not exclusive to the dataset and the problem statement. During the training phase, multiple candidate models are being developed. Each of these models is characterized by a different sets of hyperparameters in an attempt to find the best performing set of hyperparameters. In the validation phase the performance of these model is evaluated on an independent dataset: the validation dataset. Depending on the possible sets of hyperparameters, this process of training and validating is repeated multiple times until models have been trained with all possible sets of hyperparameters. The validation procedure does not prevent overfitting to the validation dataset, and therefore the performance of the selected model should be confirmed by measuring the final performance on an independent set of data. This is done with the testing dataset during the testing phase.

K-fold cross validation

For this three-phase approach, three datasets are required. The disadvantage of this method is that the part of the dataset that will be used for training is smaller than the amount of available data. To partially alleviate this issue k-fold cross-validation can be implemented. With this technique, the validation and training datasets are merged. This combined dataset is then split into k number of subsets, and the generalization loss is calculated by using $k-1$ number of subsets for training, and the remaining subset for validation. This process is visualized in Figure 3.2. Mathematically, let i denote the index of the data point and N be the total number of observations. Let $\hat{f}^{-k}(x_i)$ be the function $f(x)$, fitted the kth part of the data removed and let L be the function that calculates the loss over the predicted $\hat{y}_i$ and y_i This is done repeatedly until each subset k is used as a validation set. The overall validation loss is calculated as the average of all k validation losses, as shown in Equation 3.1[39].

$$CV(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{f}^{-k(i)}(x_i)) \tag{3.1}$$

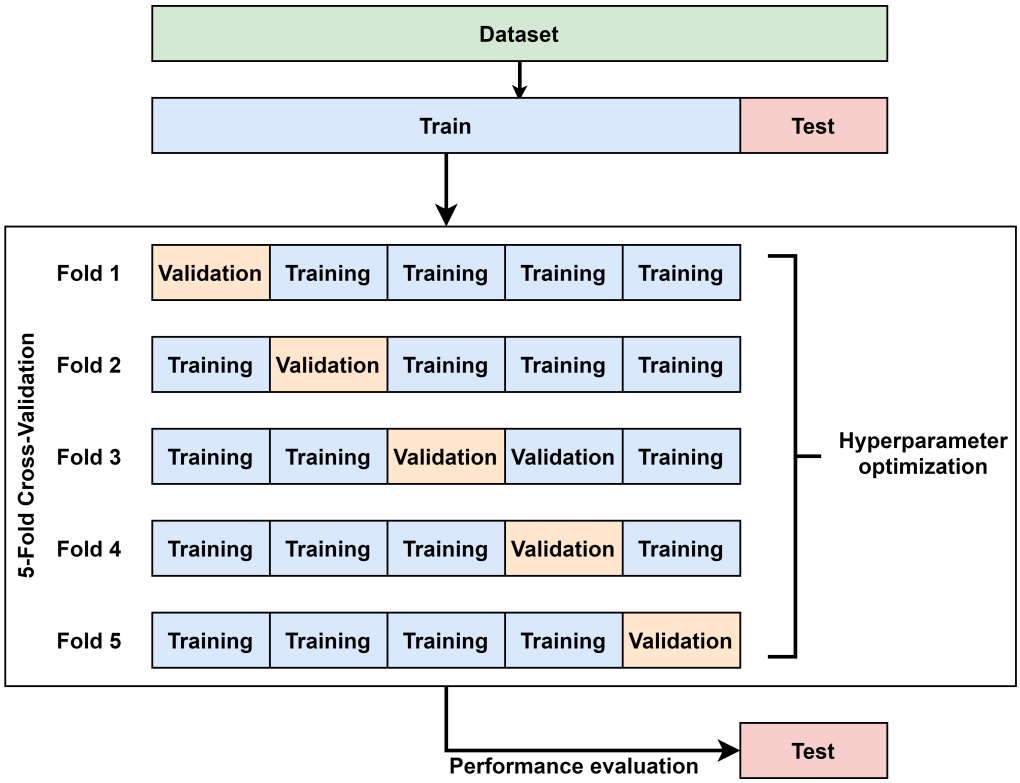

FIGURE 3.2: Visualization of the cross-validation procedure during the training process of ML models. The entire dataset is initially split into a training and testing dataset. The testing dataset is only used in the end to evaluate the performance. The training set is used to optimize hyperparameters via cross-validation

Phase 3: Testing

After models and hyperparameters have been optimized via training and validation, one candidate model remains; the model that achieved the smallest loss on the validation set. The generalization of this model is then estimated using a test dataset, a 'hold-out' dataset that was put aside for this purpose. The data from the hold-out set is never before seen by the model, but consists of data that is thought to have the same distribution as the training dataset, since all these subsets come from the original complete dataset. Ideally, the performance on this test set should be similar to the performance reached on the training set.

3.2 Introduction on ML Algorithms

The prediction of obstructive CAD is a supervised learning problem. In this learning problem, the task T can be defined as the prediction of obstructive CAD. The reference, denoted by $\mathbf{Y}$ is a binary variable that encodes if patients have obstructive CAD. The input $\mathbf{X}$ can be defined as all information, or the set of explanatory variables, that characterizes the patient. Since the outcome is binary, the set of algorithms that describe this learning problem is binary classification.

Multiple algorithms are able to perform well at binary classification tasks. Their effectiveness is strongly dependent on the complexity of the relationship between input and output, and the number of samples. It can be difficult to know how well a certain algorithm will perform on a specific classification task. Therefore I have decided to compare the performance of various algorithms that have performed well on similar classification tasks[37, 40–44]. These are logistic regression (LR), least absolute shrinkage and selection operator (LASSO), the support vector machine (SVM) and ensembled decision trees using eXtreme Gradient Boosting (XGB). Each of these models will be described to greater extent and principles will be described and visualized using an explanatory dataset.

3.2.1 Explanatory Dataset

This explanatory dataset was from Kaggle.com, a popular website for ML challenges[45]. The dataset, referred to as the Kaggle-dataset, was originally used to investigate correlations between students' alcohol consumption and students' grades. The most important features within this dataset are age, gender, math grades, weekly study time, weekly free time, alcohol consumption on weekdays and during weekends, and absence days. The entire list of features can be found in Appendix A. Instead of correlating all these features with a students' performance, I used this dataset for a binary classification problem. In terms of ML the task T can be described as the following: Predict gender based on alcohol consumption, math grades, age, study time, etc.. The models were trained for clarification purposes. For that reason, default hyperparameter optimization was used. 80% of the dataset was used for training, and 20% was used for testing. The validation phase was bypassed since there was no hyperparameter optimization done.

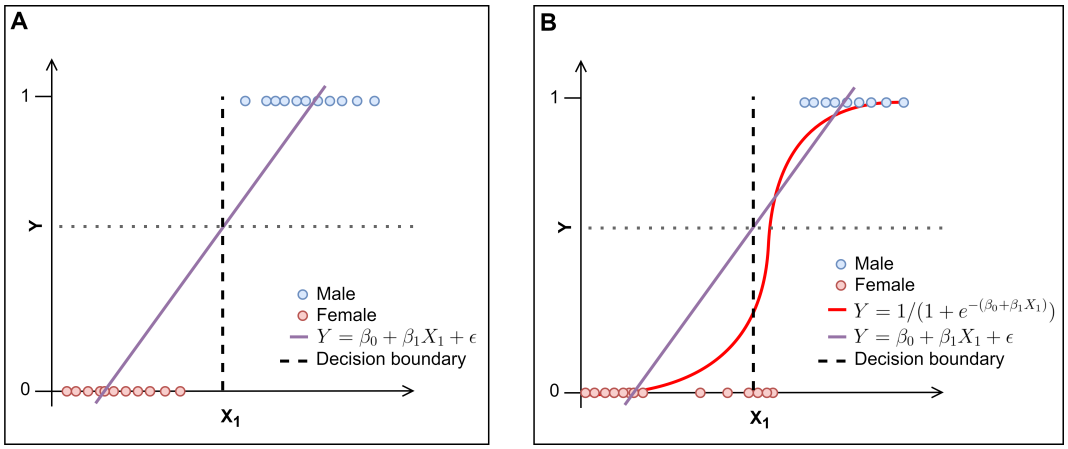

FIGURE 3.3: A) In this graph a linear relationship is modeled by means of linear regression. As per example, let X_1 encode for daily alcohol consumption in ml/day. In the Kaggle-dataset, it appears that this is a feature that can be used to discriminate gender. The decision boundary for linear regression is placed where the function $f(x)$ crosses the line $y = 0.5$. In this example, each data point is correctly classified. B) In this graph, both linear regression and logistic regression are used to model the relationship between daily alcohol consumption and gender. To highlight the weakness of linear regression for data imbalance, let's suppose that the distribution of females that drink has a skewed distribution. This is a potential cause for inaccurate classification by linear regression. In this figure this is visible when the linear classifier from figure A is compared with the linear classifier from figure B. Logistic regression is able to generate a more accurate decision boundary. An important remark is that this representation is kept one-dimensional for visualization purposes. In most cases the problem becomes high dimensional, depending on the number of features or independent variables

3.2.2 Types of Algorithms used in this Thesis

Linear Regression

A better understanding of LASSO and LR can be obtained by explaining some concepts from linear regression. Linear regression is the most basic form of modelling a relationship. Explanatory values, known as independent variables, **X** are mapped to output **Y** by an linear relationship. The number of independent variables is denoted by n. β_0 denotes the intercept, and the coefficients β_{1-n} describe the effect of a independent variable for Y. This relationship is shown in Equation 3.2 and shown in Figure 3.3A. A direct result is that the magnitude and the sign of β reveal information about the type of relationship between the corresponding feature and the reference **Y**. These coefficients are optimized by minimizing the sum of the squared residuals, also known as the mean squared error(MSE). The residuals are the difference between the Y, calculated by $f(x_i, \beta)$ and the predicted outcome $\hat{y}$, as shown in Equation 3.3.

$$f(x, \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \tag{3.2}$$

$$MSE = \sum_{i=1}^{n} (y_i - f(x_i, \beta_i))_i^2 \tag{3.3}$$

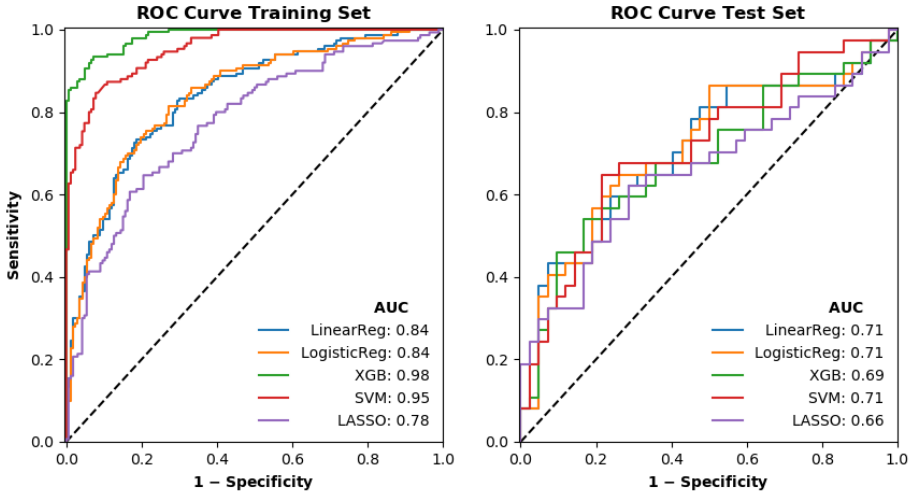

FIGURE 3.4: ROC curves and corresponding AUC values for each of the models that was trained on the Kaggle-dataset. Note that no hyperparameter optimization was used. The results on the training set show large variations, caused by overfitting of XGB and SVM, and underfitting by LASSO. The ROC curves and AUC values that correspond with the test set show that each of the models was able to discriminate gender based on the features provided in the Kaggle-dataset except. The performance by the LASSO model was notably worse than the rest.

Performance of Linear Regression on Kaggle-dataset

When linear regression was used to predict gender using the Kaggle-dataset, we find that this algorithm can be used to predict gender. Figure 3.4 shows the receiver operator characteristic (ROC) curves and the calculated area under the curve(AUC). The linear model is not a strong classifier. Nonetheless it reaches acceptable ($0.70 <$ AUC < 0.80) discriminative performance on the test dataset. The fact that the AUC is higher on the training set is an indication of overfitting. However, this does not usually translate well for problems where the outcome is categorical, for example with binary classification. One of the problems of using linear regression for classification problems is that the predicted outcome is a continuous variable, and not the probability of belonging to a class. Another issue is that it is more sensitive to data imbalance, as is illustrated in Figure 3.3. Logistic regression handles these issues and is therefore preferred for classification problems .

Logistic Regression

LR describes a statistical model that approximates the relationship between a categorical dependent variable and one or more independent variables. Although various extensions exist, the most basic form involves binary logistic regression, where the binary dependent variable has two possible outcomes, namely $Y \in \{0, 1\}$. Logistic regression attempts to model the probability belonging to a certain class. Let P be the probability for observation X of belonging to $Y = 1$, we can than take the natural logarithm of the odds and define the relationship between the odds and the independent variables, as shown in Equation 3.4. An activation function is used to convert a linear equation to the logistic regression equation. Eventually this can be simplified via Equation 3.5 to Equation 3.6. The coefficients are optimized using the maximum likelihood estimation, where the log-likelihood is maximized, and the

magnitude of the coefficients β hence indicate how much an independent variable contributes to the odds of belonging to a certain class. Figure 3.3B shows how a logistic function is able to classify data points.

$$ln(\frac{P}{1-P}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \tag{3.4}$$

$$\frac{P}{1-P} = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_n x_n} \tag{3.5}$$

$$P = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_n x_n}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_n x_n}} \tag{3.6}$$

Performance of Logistic Regression on Kaggle-dataset

We find that logistic regression performs equally well compared to linear regression (and also compared to the other algorithms). If we consider the coefficients β, it turns out that the strongest predictor for male gender in this dataset, as measured by the magnitude of β, was daily alcohol consumption. Daily alcohol consumption was a categorical variable. For that reason a visualization would not provide additional insights since we would only have a limited number of data points. However, let us suppose that daily alcohol consumption was a continuous variable measured in ml/day. In that case we could use it to make a graph similar to Figure 3.3B.

LASSO

In 1996 a new method for estimation in linear models was proposed: LASSO[46]. If we consider the estimator from linear regression, linear least squares, the estimates are calculated by minimizing the residual squared error. In the paper of Tibshirani et al. two disadvantages of this method were highlighted. Estimation using MSE usually results in low bias, but high variance, hence low prediction accuracy. It is sometimes possible to improve the accuracy by shrinking or setting some coefficients to zero. In practice, this is done by adding a so-called LASSO-penalty term to the MSE equation. Let λ denote the LASSO-penalty and let p denote the number of coefficients. The LASSO loss as a function of β and λ, denoted by $\mathcal{L}$ can then be written according to Equation 3.7 For the basic LASSo model, λ is a hyperparameter that can be altered to change the way this algorithm optimizes. The penalty reduces the variance at the cost of an additional bias, and can result in an improved overall prediction accuracy. Setting to 0 of coefficients also results in less coefficients being used in the equation. The added benefit is improved interpretability of the 'reduced' model. This is especially interesting in a clinical setting because of two reasons. Clinical implementation is often met with reservation, especially if a model acts like a black box. Providing insights in the inner workings of the model can help in the acceptation phase of implementation. Secondly, you would have to gather all required input parameters of a patient in order for a model to make a prediction. Therefore, the benefit of collecting fewer measurements while maintaining model performance can aid in the clinical implementation. LASSO is not exclusively applicable to linear or continuous problems and can be adapted to perform binary classification.

$$\mathcal{L}(\lambda, \beta) = \sum_{i=1}^{n} (y_i - f(x_i, \beta_i))_i^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{3.7}$$

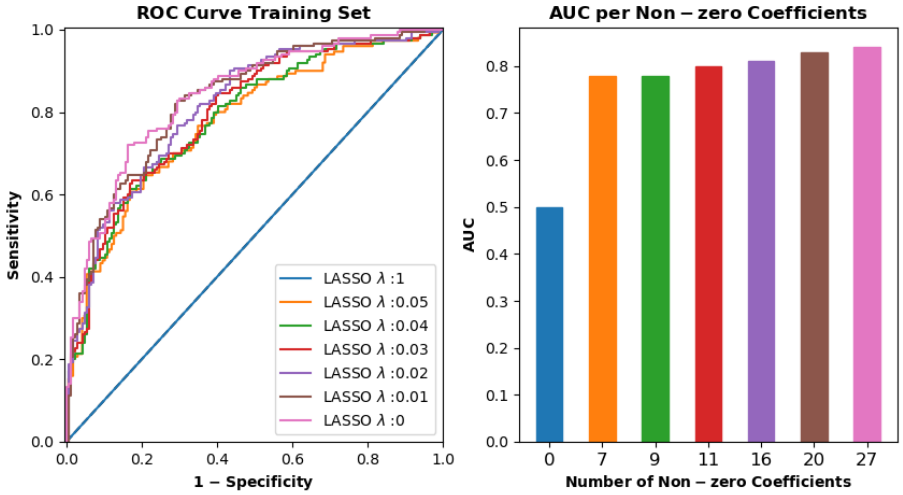

FIGURE 3.5: In this graph we can see the effect of varying the LASSO-penalty λ. When λ is set to zero, the model is identical to linear regression. When the value of λ is increased, we see a decrease in performance, but also a decrease in model complexity as measured by the number of non-zero coefficients β

Performance of LASSO on Kaggle-dataset

In Figure 3.5 the ROC curves of the LASSO classifier is graphed for various values of λ. It becomes clear that the algorithm is identical to linear regression when the LASSO-penalty is set to zero. However, when the LASSO-penalty is set too high, the coefficients will be forced zero and the prediction will be equal to the intercept, which is the mean of the dependent variable. In Figure 3.5 this is visualized: By varying the LASSO-penalty we see that the discriminative power of the model decreases. This part of the trade-off between model complexity and model performance. When the performance (AUC) is graphed against the number of non-zero coefficients β we see that even though the performance decreases, so does the number of coefficients (e.g. features) that is used in the model. Only a slight decrease in performance is achieved with just seven features, as opposed to using all 27 features with linear regression.

Support Vector Machine

The SVM algorithm is well established as a binary classifier[47]. The SVM generates a model that represents the data as points in space, and within this space a decision boundary or hyperplane is then determined. This principle is shown in Figure 3.6. This hyperplane is constructed using the support vectors: data points that are most prone to misclassification. The hyperplane is constructed in the point space as to maximize the margin between data points of both classes. Unseen examples are classified based on which side of the hyperplane they are positioned in the point space.

Performance of SVM on Kaggle-dataset

Figure 3.4 shows the performance of a SVM model on the Kaggle-dataset. From this figure we can learn a number of things. First of all, this is the first algorithm presenting a clear case of overfitting. We find a significant discrepancy between the

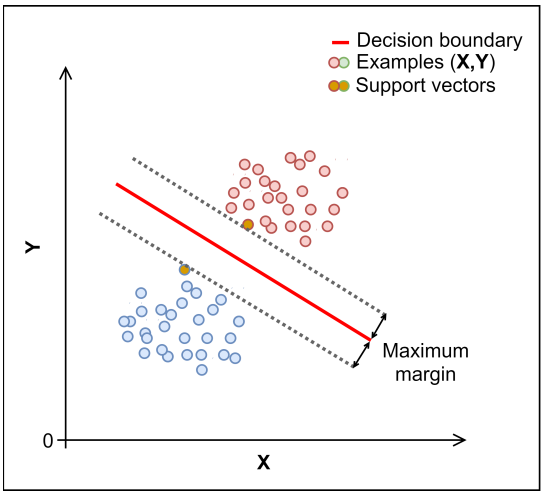

FIGURE 3.6: After initial transformation of the data points to a point space, support Vectors are formed by data points that lie the closest together. This figure shows a how a hypothetical decision boundary is determined that maximizes the margins between the support vectors, marked with a yellow center.

performance of the model on the training dataset and the performance on the test dataset. No hyperparameter optimalization was done for the development of this model. The model that was used for this figure is a clear example of when regularization measures should be implemented to reduce overfitting. Another idea that we can take from this figure stems from the fact that the model was able to overfit. As opposed to the simpler algorithms such as linear regression, LASSO and Logistic regression, the SVM was able to utilize more relationships within the data. This does not necessarily mean that the simpler algorithms can not be altered via hyperparameters to fit more complicated data, but it does show that a SVM is capable of fitting high dimensional data. Arguably, the relationships that are found by this SVM model are probably not useful for the prediction of gender, and therefore only lead to overfitting. Since data is transformed to point space, it becomes difficult to identify the contribution of single features to the final prediction, which can be interpreted as a disadvantage.

Decision Trees

A decision tree(DT) as a ML algorithm can be used for both regression and classification. A basic DT is shown in Figure 9. The DT models a pathway of different outcomes, where 'decisions' along the way lead to a specific outcome. It can be dissected into branches and leaves. The branches split into other branches at a node, and the node represents a decision. The decision determines which branch to follow. Eventually the branch will end and reach an endpoint. These endpoints are called leaves, and encode for a certain class. The learning occurs iteratively by partitioning the data by forcing decisions. The nodes are updated each iteration to optimize with regards to the performance metric. The relative simplicity and intuitive interpretability of decision trees have popularized their usage, as well as their excellent performance for certain tasks[48]. complexity of the DT is dependent of the available data, but can be controlled with parameters that define the depth, e.g. the number of

decisions to reach a class. DT can model non-linear relationships, but are also prone to overfitting, and regularization measures are often required to prevent this.

Learning Ensembles of Functions

Another approach to solving difficult problems statements with decision trees whilst maintaining relative simplicity and preventing overfitting is to use multiple DT's. Individual DT are then regarded as 'weak learners'. An ensemble of DT can be created to forms a strong 'learner'. The prediction of an ensemble model is then determined by the combined votes of each weak learner. This basic idea is shown in Figure 9B Two common methods to develop ensembles are bootstrap aggregating (bagging) and boosting. In bootstrap aggregating each model has the same voting weight, but each model is trained with a random subset of the training dataset to increase the variance of models. This can be combined with a random selection of DT parameters to generate random forests. With boosting, each model is trained with the aim to correctly classify the misclassified cases of previous models. This is done by attributing extra weight to these previously misclassified cases.

Boosting can be combined with gradient descent techniques to efficiently optimize parameters of models[49]. These so called gradient boosted decision trees are currently well established classification algorithms, having outperformed other approaches including SVM, Naïve Bayes and random forests in a recent comparison on 71 datasets[50]. The XGB algorithm was used in this thesis.

Performance of XGB on Kaggle-dataset

The results of the XGBoost model on the Kaggle-dataset can be interpreted in a similar manner as the SVM model. This is another case of overfitting, but once more, the tendency of the model to overfit indicates that it is able to utilize complicated, nonlinear correlations. With boosted decision trees it is easier to determine the importance of features. We can simply derive how often certain features occur in nodes, and determine the decrease in model performance if a feature was removed from the dataset. The magnitude of decrease in model performance then inversely encodes for the feature importance. It is also possible to plot singular trees, an example is provided in Figure 3.7.

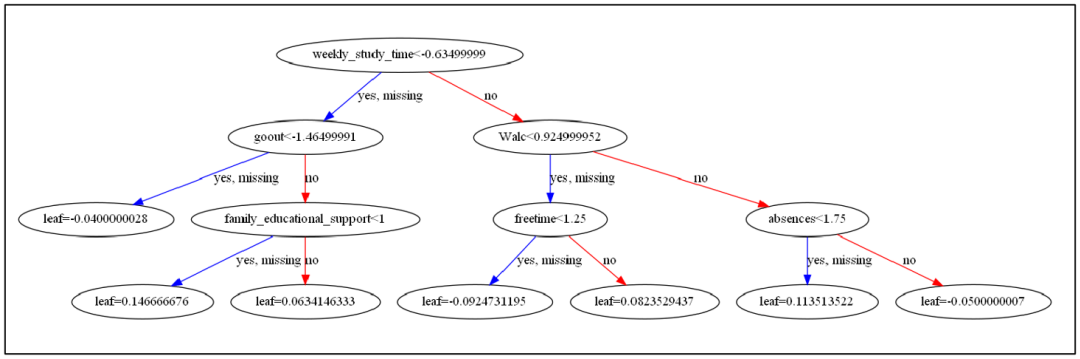

FIGURE 3.7: One out of 100 decision trees used in the model that was trained on the Kaggle dataset. Via decisions in the nodes, an endpoint that encodes for the probability of belonging to the class for male gender, is reached. The final probability is the summation of the probabilities of all trees combined.

Chapter 4

Prediction of Obstructive Coronary Artery Disease using Machine Learning Algorithms

4.1 Introduction

Cardiovascular disease is the number one cause of death globally[51]. In Europe, it is the single largest contributor to mortality, accounting for 40% of all deaths. Moreover, 19% of all deaths in the European Union were due to coronary artery disease (CAD)[52] . Furthermore, despite the recent decreasing trend in mortality due to CAD, the burden of CAD is not confined by mortality because the average hospitalization duration has increased[52].

The latest European Society of Cardiology (ESC) guidelines propose a six-step approach for the management of patients with angina and suspected CAD[9]. After symptom investigation, consideration of comorbidities and basic testing, the pre-test probability(PTP) and clinical likelihood of CAD is determined. The PTP can be estimated by clinical models. State of the art PTP models currently incorporate part of the accumulation of data that has been gathered of a patient, including patient characteristics, risk factors, ECG changes, coronary artery calcification[9]. Additional examinations such as CTA and PET MPI provide additional insights. Myocardial Perfusion Imaging (MPI) using positron emission tomography (PET) or single photon emission tomography (SPECT) are commonly used non-invasive techniques to evaluate the myocardial perfusion. The quantification of the myocardial blood flow (MBF) and myocardial flow reserve (MFR) further improves the diagnostic accuracy for detecting significant CAD[23, 53].

The amount of available data that needs to be interpreted for diagnosis keeps growing. This data may be utilized to improve diagnostic accuracy. However, the fact that some of these variables intercorrelate with complicated relationships as well as the shear amount of variables make it increasingly more difficult to interpret in clinical practice. This causes risk stratification of patients with suspected obstructive CAD to remain a challenging task. Recent studies have attempted to improve the identification of obstructive CAD by adapting existing predictive models. They have shown the potential of integrating imaging derived features with clinical data in risk prediction models for improved risk stratification[35–37]. In fact, machine learning (ML) algorithms are excellent for this task because of their ability to establish relationships between these features and patient outcome, regardless of the

number of variables.

Some of these ML approaches have already shown the benefit of the integration of different data types for the detection of significant CAD. Juarez-Orozco et al. developed several ML models to predict obstructive CAD as determined by SPECT and concluded that ML is a feasible and applicable method to identify patients who will present CAD[37]. Reeh et al. developed a model which expanded on the modified Diamond Forrester model with additional clinical information, resulting in improved identification of low risk subgroups[41]. Furthermore, Hu et al. showed that deep learning methods that incorporated clinical data and SPECT imaging for the prediction of obstructive disease improved the automatic prediction of CAD compared to the current standard of care[42].

PET MPI is a well established technqiue for cardiac blood flow quantification[32]. We hypothesize that parameters derived from PET MPI can be used more extensively in the diagnostic pathway. Hence, our aim is to derive and test ML algorithms to obtain an individual risk stratification of obstructive CAD after PET MPI and CT coronary artery calcification(CAC) scoring and compare this with the diagnostic accuracy attained by clinicians.

4.2 Methods

4.2.1 Study Design

We retrospectively included 1007 consecutive patients with suspected CAD. These patients had no prior history of CAD and underwent a CAC-score, and were referred for rest and Regadenoson-induced stress Rubidium-82 PET/CT (Discovery 690, GE Healthcare). Cardiac risk factors; cigarette smoking, hypertension, hypercholesterolemia, diabetes, positive family history of CAD; prior medical history; age; gender; body mass index (BMI); creatinine serum values; coronary artery calcification (CAC) score and medication usage were registered at time of the PET/CT examination. Patients were classified as having obstructive CAD if follow-up included either a conclusive invasive coronary angiography (ICA) for CAD as defined by a significant FFR measurement (< 0.8) or >70% stenosis on ICA or a revascularization during follow-up including percutaneous coronary intervention (PCI) or Coronary Artery Bypass Grafting (CABG) procedure. Obstructive events were retrieved from electronic patient records whilst maintaining a minimum follow-up time of 1 year.

4.2.2 MPI Data Acquisition and Reconstruction

Prior to MPI, a low-dose CT scan was acquired during free-breathing to provide an attenuation map of the chest. This scan was made using a 5-mm slice thickness, 0.8 s rotation time, pitch of 0.97, collimation of 32x0.625 mm, tube voltage of 120 kV, and a tube current of 10 mA. Next, 740 MBq Rb-82 was administered intravenously with a flow rate of 50 mL/min using a Sr-82/Rb-82 generator (CardioGen-82, Bracco Diagnostics Inc.). After the first elution, we induced pharmacological stress by administrating 400 g (5 mL) of Regadenoson over 10 seconds. After a 5 mL saline flush (NaCl 0.9%), we administered a second dose of 740 MBq Rb-82. Seven-minute PET list-mode acquisitions were acquired after both Rb-82 administrations. Attenuation

correction was applied to all data on the PET system after manual registration of CT and PET data.

The acquisitions were reconstructed using 26 time frames (12x5 s, 6x10 s, 4x20 s and 4x40 s) with default settings as recommended by the manufacturer using 3D iterative reconstruction using 2 iterations and 24 subsets, while correcting for decay, attenuation, scatter and random coincidences, and dead time effects. Neither time-of-flight information or resolution modeling was used.

CT-based CAC-score scans were performed using a 64-slice CT scanner (LightSpeed VCT XT, GE Healthcare, StataSE 12.0, StataCorp LP, College Station, Texas, USA). An unenhanced ECG-gated scan was obtained prospectively, triggered at 75% of the R–R interval by using the following scanning parameters: 2.5mm slice thickness; gantry rotation time, 330ms; tube voltage, 120 kV; and a tube current of 125–250mA, depending on patients' size.

4.2.3 Data processing

The reconstructed dynamic scans were post-processed using Corridor4DM software (v2015.02.64). Myocardium contours were automatically detected in both rest and stress scans based on the dynamic images. Furthermore, a region of interest (ROI) was manually placed at the location of the mitral valve to estimate the activity in the blood pool. The activity concentrations in the myocardium contour and ROI were measured in the 26 reconstructed time frames to calculate the time activity curves (TACs) for the left ventricle (LV), for the three vascular territories: left anterior descending (LAD), left circumflex (LCX) and right coronary (RCA) artery, and for the whole myocardium. The left ventricular ejection fraction(EF) in rest and in stress was automatically generated. The one-tissue compartment model of Lortie et al. based on a ROI methodology was used to calculate the MBF[31]. The MBF was calculated for each of the three major coronary arteries for both stress and rest phases from the TACs using Corridor4DM. The global MBF values were calculated by averaging the MBF of all three vessels. The MFR was derived from the MBF as the ratio between the MBF in stress and the MBF in rest[31]. Summed Stress Scores(SSS) and Summed Difference Scores(SDS) were derived automatically within Corridor4DM and are semi-quantitative evaluations of the perfusion of the myocardium. The SSS was derived from the stress image and reflects the presence, extent and severity of perfusion defects, whilst the SDS is the difference between the SSS and the same semi-quantitative evaluation during the rest phase. The SDS indicates the reversibility of a defect.

Myocardial perfusion entropy(MPE) is suggested to provide prognostic information for the occurrence of major adversarial cardiac events (MACE)[54]. Entropy can be understood as a measure of order. An analogy can be made by looking at water. When water is frozen, the molecules are arranged in a somewhat orderly manner and the entropy of the system is relatively low. If we compare ice with water, we find that non-frozen water has a higher level of entropy. In human phsyiology, entropy is a metric for quantifying the irregularity and/or complexity contained in signals[55]. MPE can be understood as the amount of disorder within the myocardial perfusion.

Since the regional flow values (of each of the 17 segments) were available for data analysis, we decided to calculate the MPE from the regional myocardial flow

reserve (MFR) values using Shannon's equation for entropy, as seen in Equation 4.1, multiplied with the average entropy of the 17 segments[56]. In this equation, $p(x)_i$ is the probability of state i.

$$H(X) = -\sum_{i=1}^{n} p(x_i) log_2 p(x_i) \tag{4.1}$$

Dedicated software (SmartScore, Advantage Windows 4.4, GE Healthcare) was used to calculate the CAC score per vessel according to the standard Agatston criteria, in which the area of calcification (in mm2) of pixels with an intensity larger than 130 HU is multiplied with a factor depending on density[12]. Calcifications were manually assigned to the vessels by experienced operators to ensure inclusion of all calcified regions, accurate allocation of the coronary arteries, and to exclude calcium deposits outside the coronary arteries.

4.2.4 Clinical Evaluation

The clinical diagnosis was determined by the evaluation by a team of two clinicians (a cardiologist and a nuclear physician) of the clinical report and the MPI scan. A positive diagnosis for obstructive CAD was defined as a consensus of possible or definite ischemia on the MPI scan by a nuclear medicine clinician and a cardiologist. This clinical evaluation parameter is referred to as the visual interpretation of the scan. Three levels of visual interpretation can be distinguished: 1) no ischemia, 2) possible ischemia and 3) ischemia.

4.2.5 Machine Learning

We have implemented several machine learning algorithms from the Scikit-Learn library in python for binary classification of the presence of an obstructive CAD event; the Least Absolute Shrinkage and Selection Operator (LASSO), Logistic Regression (LR), Support Vector Machine (SVM) and the XGBoost implementation of gradient boosted decision trees[57].

Every patient was characterized by an array of features, including PET-MPI and CAC findings, and the various other clinical features shown in Table III. All features except CAC score, age, resting heart rate and BMI were transformed to dichotomous variables. The remaining continuous variables were scaled by standardization. Subsequently, the dataset was randomly split into a train and test set with a 4:1 ratio, stratified by occurrence of obstructive events, so that both the train and test sets had a comparable prevalence of obstructive CAD. Subsequently we evaluated a hybrid approach where the visual interpretation by the clinicians was used as an additional input parameter. Each of the algorithms was optimized once more whilst including this additional input parameter.

Model Development

The ML algorithms were optimized with the training dataset, using hyperparameter optimization via grid search in combination with 3-times repeated, 5-fold stratified

cross validation. Hyperparameters are parameters that control the learning process of an algorithm. Different values and different combinations of hyperparameters can result in alterations in model performance, and the correct set of hyperparameters is often specific to the dataset that is used. Therefore, various combinations of hyperparameters should be tried. A grid search refers to the method of combining various hyperparameters while evaluating model performance. The grid search table in Table 4.2 shows the different hyperparameters per algorithm. For the SVM, several regularization parameters were used as well some kernel functions. For XGBoost, the hyperparamters that controlled the model complexity included depth, and number of estimators. The most important regularization parameters for XGBoost were is the fraction of features that was used to train each tree, known as 'Colsample by tree' and the fraction of training data that was used to train each tree, known as the subsample ratio. For LASSO there is only the alpha value to modify the LASSO penalty. Model performance was evaluated using the F1-score and can be derived from Equations 4.2-4.4. This metric can be understood as the harmonic mean of precision and sensitivity especially useful in learning problems with an imbalanced dataset.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{4.2}$$

$$Sensitivity = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{4.3}$$

$$F1 - Score = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity} \tag{4.4}$$

For each algorithm, the best performing model was saved and the performance was evaluated on the test set. It is clinically relevant which features a model uses as input since this can aid the clinicans understanding of the model. Therefore, for the XGBoost model, the order of feature importance was extracted from the model. This feature importance ranking is based on the number of times a feature appeared in decision trees within the particular model.

4.2.6 Statistical Analysis

IBM SPSS (IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp) was used for statistical analysis. Differences in patient characteristics between the training and test dataset were evaluated using Student's t-test, if the characteristic was distributed normally according to the Shaprio Wilk test, else the Mann-Whitney U test was used after validation of homogeneity of variance using Levene's Statistic. Bonferroni correction for multiple testing was applied to test if the two groups were significantly different from each other[58, 59]. For each ML algorithm, the area under the receiver operating curve (AUC), the sensitivity, specificity, accuracy and the F1-score were calculated. The comparison of the performance of the different models with the ability of clinicians to diagnose obstructive CAD was made by looking at sensitivity, specificity, accuracy and the F1-Score. Confidence intervals for sensitivity and specificity were calculated using the exact Clopper-Pearson method[60]. A significance level at 0.05 was used for all statistical tests. This resulted in a p-value of 0.0013 after Bonferroni correction. Furthermore the predictive performance

of the ML algorithms was compared with the diagnostic performance by clinicians by comparing the sensitivity, specificity and the F1-Score.

4.3 Results

4.3.1 Study Population

Patient characteristics and clinical parameters of the study population are shown in Table 4.1. Within the total population of 1007 patients there were 111 patients with obstructive CAD (11%) during follow-up. The average time for follow-up was 1.8 years. The minimum follow-up time was 1 year whilst the longest follow-up time was 3.8 years. 26 deaths were reported from this study population. The cause of death was not always reported. The majority of obstructive events occurred within 90 days after the PET scan, and can be referred to as early revascularization. 67,4% of the obstructive events in the training dataset and 72,7% of the obstructive events in the test dataset can be thought of as early revascularization.

A number of significant differences were found between the characteristics of patients in the training dataset compared with patients in the test dataset. The percentage of patients with a medical history was larger in the test dataset, namely 26.7% versus 19.5% in the training dataset (p=0.04), and so was the share of patients in the test set that used beta blockers, namely 62.4% versus 53.8% (p=0.03). Furthermore, the distribution of the SDS score within the training dataset showed dissimilarities compared to the test dataset. However, after correction for multiple testing using the Bonferroni correction, we state that the baseline characteristics of the two groups as a whole did not differ significantly from each other.

TABLE 4.1: Comparison of patient characteristics and all features between the training and test dataset. Normally distributed continuous features are characterized by mean values and corresponding standard deviations. Non-normally distributed continuous features are denoted by *, and are characterized by median values and the corresponding interquartile range. Dichotomous features are characterized by the percentage of occurrences.

Characteristic	**Training set** (n = 805)	**Test set** (n = 202)	**p-value**
Obstructive CAD	11.1% (89)	10.9% (22)	0.95
All Cause Death	2.6% (21)	2.5% (5)	0.91
Time to event* (days)	42 (28-153)	34 (26.5-84.5)	0.89
Event <90 days after scan	67,4% (60)	72,7% (16)	0.82
Age	66.2 ± 10.6	65.7 ± 10.9	0.54
Female	49.70%	48.50%	0.77
Length (cm)	173.6 ± 10.3	174.1 ± 10.5	0.54
Weight (kg)	89.0 ± 19.9	89.4 ± 19.8	0.80
BMI	25.6 ± 5.2	25.6 ± 5.14	0.93
Pulse (beats/min)	69.8 ± 13.6	70.4 ± 11.3	0.56
Creatinine (µmol/L)	97.2 ± 75.9	89.9 ± 40.7	0.19
Smoking never	40.1%	38.60%	0.70
Smoking ever	47.3%	46.5%	0.84
Smoking present	12.5%	14.8%	0.38
Diabetes mellitus	20.0%	17.8%	0.49
Hypercholesterolemia	41.1%	40.6%	0.89
Hypertension	62.3%	61.1%	0.80
Positive Family history	51.1%	46.5%	0.25
Medical history	19.5%	26.7%	0.04
COPD	12.0%	16.3%	0.12
CVA	9.2%	12.9%	0.15
Medication usage	94.3%	97.0%	0.06
Aspirin	28.2%	27.2%	0.78
Clopidogrel	4.1%	4.5%	0.82
Acenocoumerol	9.0%	10.9%	0.40
Beta blockage	53.8%	62.4%	0.03
Ace/AII inhibitor	41.4%	45.0%	0.35
Ca-channel blocker	24.2%	18.3%	0.06
Statin	43.2%	47.0%	0.33
Diuretic	28.9%	35.6%	0.07
Agatston score* (Total)	120 (10-551)	153 (15-663)	0.31
Agatston score* (LM)	0 (0-9)	0 (0-8)	0.93
Agatston score* (LAD)	56 (1-251)	94 (1-270)	0.29
Agatston score* (LCX)	4 (0-79)	3 (0-74)	0.54
Agatston score* (RCA)	5 (0-110)	9 (0-190)	0.17
PET SSS*	4 (1-9)	4 (1-10)	0.47
PET SDS*	1 (0-3)	1 (0-3)	0.02
MBF Stress* (ml/min/g)	2.5 ± 0.7	2.5 ± 0.8	0.30
MBF Rest* (ml/min/g)	1.1 ± 0.3	1.1 ± 0.3	0.75
MBF Reserve* (ml/min/g)	2.5 ± 0.6	2.4 ± 0.6	0.36

TABLE 4.2: This table shows the values of hyperparameter that were used throughout the grid search per algorithm. Four algorithms were used: Support Vector Machine (SVM), eXtreme Gradient Boosting (XGB), Least Absolute Shrinkage and Selection Operator (LASSO) and Logistic Regression(LR). Sets of hyperparameters that resulted in overall best performance are color-coded. Hyperparameters in red indicate that the hyperparameter was selected in the optimization of the regular models. Hyperparameters in blue indicate that the hyperparameter was selected in the optimization in the hybrid models. Hyperparameters in green indicate that the hyperparameter was selected in both models.

Classifier	**Parameters**	**Range or domain**
SVM	Kernel	[Sigmoid, Radial Basis Function]
	γ	[1e-5, 1e-4, 1e-3, 1e-2, 0.1, 1, 10]
	C	[1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 10, 25, 50, 100, 500]
	Shrinking	[True, False]
XGB	Scale Positive Weight	[1, 3, 5, 7, 9, 11]
	Colsample by Tree Ratio	[0.5, 0.6, 0.7, 0.8, 0.9]
	Subsample Ratio	[0.5, 0.6, 0.7, 0.8, 0.9]
	γ	[0, 0.1, 0.2, 0.3, 0.4, 0.5]
	Learning Rate	[0.1, 0.01, 0.05]
	Max Depth	[3, 5, 7, 9]
	Number of Estimators	[50, 100, 150, 200]
LASSO	α	[1e-5, 5e-4, 1e-4, 5e-4, 1e-3, 5e-3, 1e-2, 5e-2, 0.1, 0.5, 1, 5, 10, 50, 100, 500]
LR	Penalty	[L1, L2, None]
	C	[0.01, 0.1, 1, 10, 100, 1000]
	Class Weight	[Balanced, None]
	L1 Ratio	[0.1, 0.5, 0.01]

4.3.2 Model Development

Four ML algorithms were trained for the prediction of obstructive CAD using 5-fold cross validation in combination with a grid search. To clarify: this process was repeated twice; the first iteration involved all input features except the visual interpretation by clinicians, referred to as the 'regular approach'. The second iteration involved all input features, and is referred to as the 'hybrid approach'. The resulting sets of hyperparameters, for each model and for both approaches, are shown in Table 4.2.

ROC curves were generated for all models to evaluate their performance on the training and test dataset. The performance of the regular models on the training dataset is shown in Figure 4.1 by means of ROC and AUC. The best performing model during cross validation was XGB which reached an AUC of 0.92, the SVM performed the worsts, obtaining an AUC of 0.76. A visual comparison of performance metrics is provided in figure 4.3. The exact numbers, including confidence intervals, are given in Table 4.3.

TABLE 4.3: Regular model performance on training dataset

	Sensitivity	**95% CI**	**Specificity**	**95% CI**	**F1-Score**
Clinician	55.1%	44.1-65.6	93.4%	91.4-95.1	0.53
XGB	57.3%	46.4-67.7	96.9%	95.4-98.1	0.63
LR	41.6%	31.2-52.5	98.5%	97.3-99.2	0.54
LASSO	21.4%	13.4-31.3	99.3%	98.4-99.8	0.34
SVM	33.7%	24.0-44.5	96.8%	95.2-98.0	0.42

TABLE 4.4: Hybrid model performance on training dataset

	Sensitivity	**95% CI**	**Specificity**	**95% CI**	**F1-Score**
Clinician	55.1%	44.1-65.6	93.4%	91.4-95.1	0.53
XGB	65.2%	54.3-75.0	97.4%	95.9-98.4	0.70
LR	49.4%	38.7-60.3	98.6%	97.5-99.3	0.62
LASSO	28.1%	19.1-38.6	99.3%	98.4-99.8	0.42
SVM	59.6%	48.6-69.8	100.0%	99.5-100.0	0.75

The performance of the hybrid models on the training dataset is shown in Figure 4.2 by means of ROC and AUC. The best performing model during cross validation was SVM with an AUC of 0.98. The worst performing model was LASSO with an AUC of 0.89. Again, the visual comparison of performance metrics is provided in Figure 4.4. The exact numbers including confidence intervals are shown in Table 4.4.

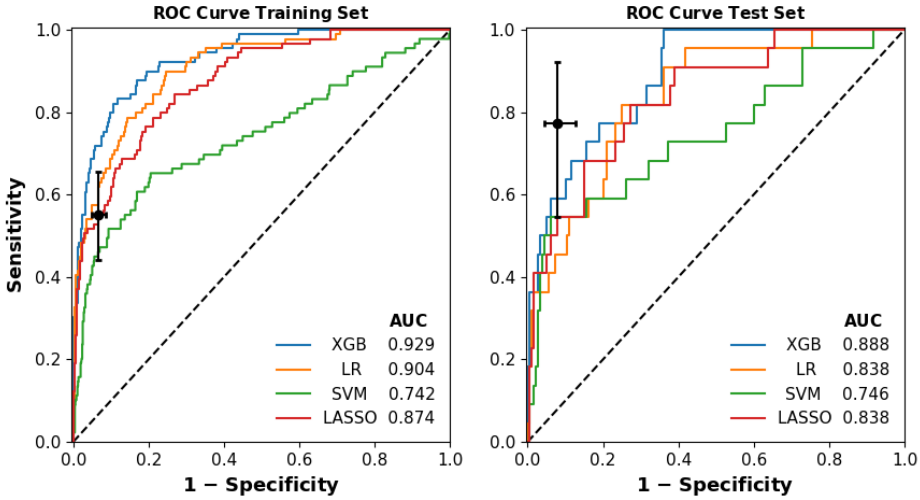

FIGURE 4.1: ROC curves of the best performing regular models per algorithm on the training and test dataset respectively. The sensitivity and specificity of the clinicians, as measured by visual interpretation of PET MPI scan, is plotted with corresponding 95% confidence intervals.

TABLE 4.5: Regular model performance on test dataset

	Sensitivity	95% CI	Specificity	95% CI	F1-Score
Clinician	77.3%	54.6-92.2	92.2%	87.3-95.7	0.64
XGB	54.6%	32.2-75.6	94.4%	90.0-97.5	0.54
LR	36.4%	17.2-59.3	95.0%	90.7-97.7	0.41
LASSO	22.7%	7.82-45.4	98.9%	96.0-99.9	0.44
SVM	36.4%	17.2-59.3	96.7%	92.9-98.8	0.37

4.3.3 Model evaluation

The four best performing models for both approaches (regular and hybrid) were evaluated using the test dataset.

The performance of the regular models on the test dataset is shown in Figure 4.1. The best performing algorithm was XGBoost which reached an AUC of 0.90. Whilst using the standard threshold of 0.5, this resulted in a sensitivity of 55% (95% CI 32-76) and a specificity of 94% (95% CI 90% to 97%). Performance metrics of all other models are shown in Figure 4.3 and the corresponding values are provided in Table 4.5.

The performance of the hybrid models on the test dataset is shown in Figure 4.2. The best performing algorithm was XGBoost which reached an AUC of 0.93. The sensitivity of this model was 64% (95% CI 41-83) and the specificity was 96% (95% CI 91-98) Performance metrics of all other models are provided in Table 4.6 and a visual interpretation is available in Figure 4.4.

TABLE 4.6: Hybrid model performance on test dataset

	Sensitivity	95% CI	Specificity	95% CI	F1-Score
Clinician	77.3%	54.6-92.2	92.2%	87.3-95.7	0.64
XGB	63.6%	40.7-82.8	95.6%	91.4-98.1	0.64
LR	40.9%	20.7-63.7	93.9%	89.3-96.9	0.43
LASSO	45.5%	24.4-67.8	98.3%	95.2-99.7	0.57
SVM	22.7%	7.8-45.4	98.3%	95.2-99.7	0.37

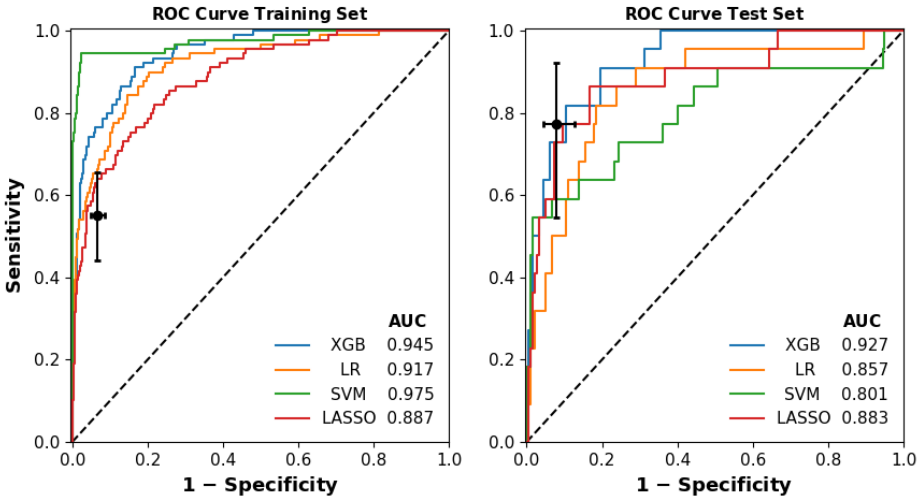

FIGURE 4.2: ROC curves of the best performing hybrid models per algorithm on the training and test dataset respectively.The sensitivity and specificity of the clinicians, as measured by visual interpretation of PET MPI scan, is plotted with corresponding 95% confidence intervals. The SVM model shows clear overfitting in the training dataset, and a lack of ability to perform on the test dataset.

Features were ranked in order of importance for both the regular and hybrid XGBoost models. The top 10 predictors for the XGBoost model consisted of PET derived features (in order of importance: PET SDS, MPE. PET SSS, MFR LAD, MBF Stress LAD), CAC-scoring (total, RCX, LAD, RCA) and creatinine serum level. The feature importance for the regular XGBoost model can be examined in greater detail in Figure 4.5, and the feature importance for the hybrid model is shown in figure 4.6.

Features that had a feature importance of zero had negligible little to none predictive value for obstructive CAD. There were only slight variations between the hybrid and the regular model. Noteworthy, features that did not contribute to both models were COPD, past CVA, hypertension, medication usage and clopidogrel usage. All prescribed medication categories were found to hold little to none predictive value.

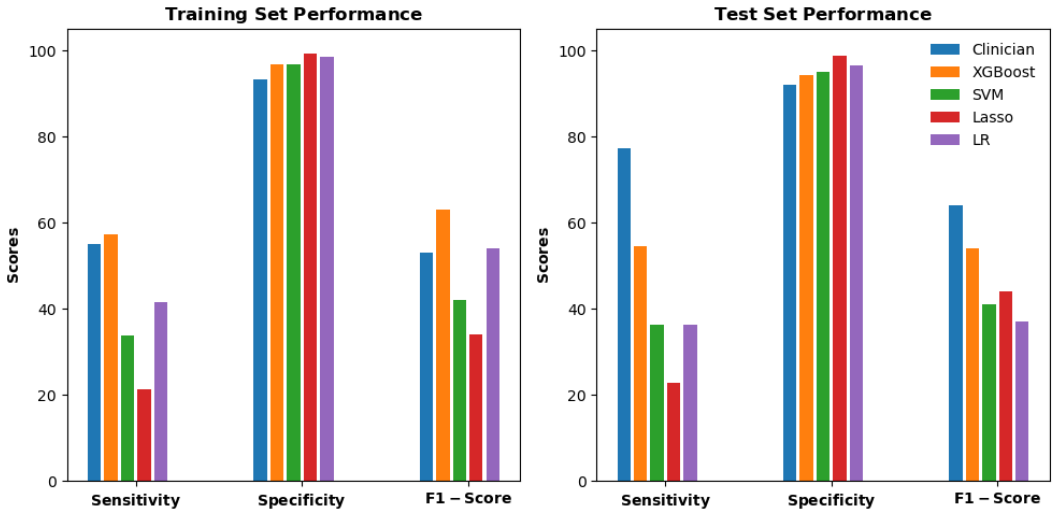

FIGURE 4.3: Visual representation of regular model performance on the training dataset and the test dataset.

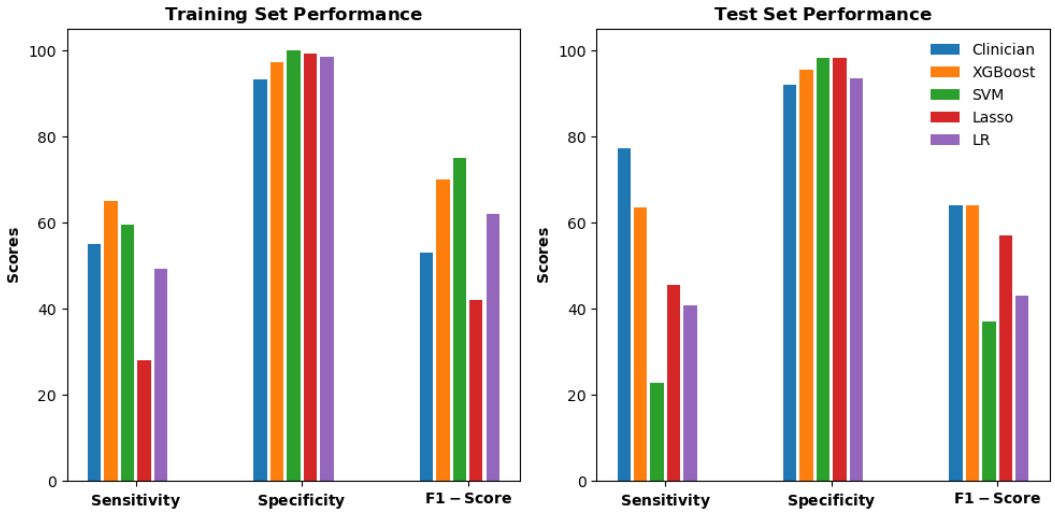

FIGURE 4.4: Hybrid model performance on the training dataset and the test dataset.

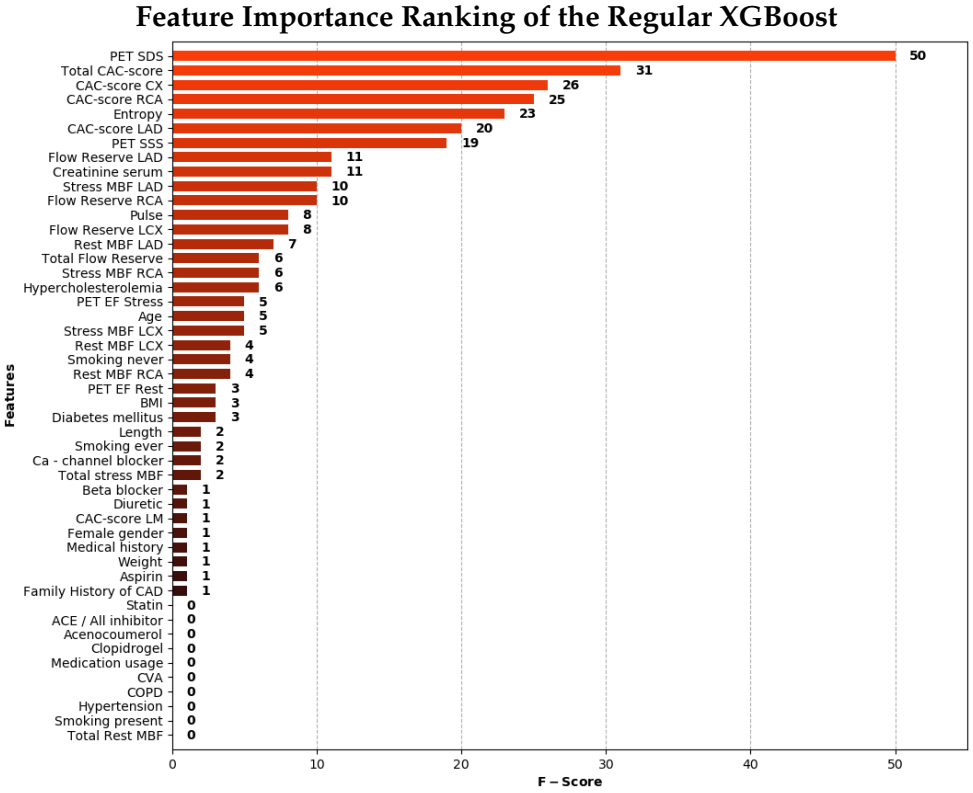

FIGURE 4.5: Feature importance ranking of the regular XGBoost model. The F-score was calculated by the improvement in accuracy brought by a feature to the branches it is on. Features with low importance can be interpreted as weak predictors for obstructive CAD

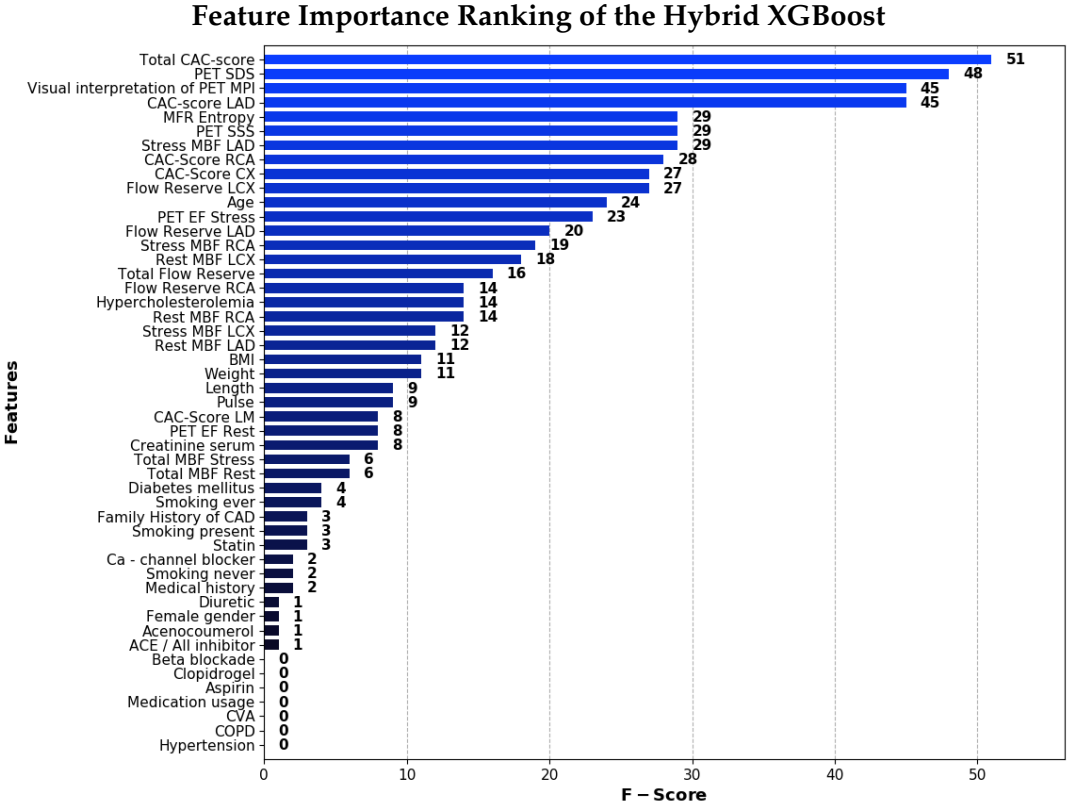

FIGURE 4.6: Feature importance ranking of the hybrid XGBoost model.

4.4 Discussion

In this study, ML algorithms were used to predict obstructive CAD based on a combination of clinical parameters and imaging derived quantitative features. We used a unique combination of classical risk factors, patient characteristics and comprehensive imaging derived features including perfusion, cardiac function and the CAC-score. We were able to develop a boosted decision tree ensemble model that provides an individualized risk score for obstructive CAD. Furthermore, we were able to identify the rationale of the model by analyzing the feature importance. This may help clinicians understand the model, but as an added benefit, this may also be used to identify previously unknown correlations. The regular model reached an AUC of 0.90 while the hybrid model reached an AUC of 0.93. It is noteworthy that this level of performance is unmatched by other approaches to predict cardiovascular risk[37, 40, 42].

On the training dataset some of the models showed similar performance as the clinician, however this did not translate towards the test dataset. All models were indicative of some overfitting, however the SVM model performed worse compared to the other algorithms, both on the training and test datasets. This is highlighted by the relative low AUC values. The model that was able to achieve the highest AUC value was XGBoost. This model performed similar to the evaluation of the scan by a clinician, but the relatively low number of positive cases in the test dataset (n=22) resulted in large confidence intervals, making it impossible to show a significant difference in comparison with clinical performance.

No models showed significant improvements after the addition of the visual interpretation, though during this cross-validation procedure, the SVM model did overfit on the training dataset, more so than the non-hybrid model. The slight improvements were expected since the visual interpretation is rarely just an objective evaluation of the scan alone. The clinician analyses the image, and places the outcome of the scan in the context of a patient. The interpretation is therefore guided by experience, but also by everything else that is known of a specific patient, such as symptoms, biochemistry results and the CAC-score.

Again, XGBoost achieved the highest AUC of all algorithms. This comes as no surprise as boosted gradient decision trees have surpassed other algorithmic classification approaches in recent years[50]. We were also able to identify the most important predictors via feature importance ranking of the XGBoost model. PET derived quantification values including MPE, SDS and SSS, as well as CAC-scores were much stronger predictors for obstructive CAD compared to classical risk factors such as smoking and hypertension. This is not unexpected, since both CAC-scoring and MPI findings are well established as independent and complementary predictors[11, 19, 21, 61, 62]. Furthermore, the creatinine serum level was ranked as a relatively strong predictor. This may be explained by the experience that renal dysfunction increases the likelihood of CAD and has a negative impact on the prognosis[63].

An interesting finding is that MPE was ranked as a strong predictor for obstructive CAD. MPE seems to provide additional prognostic information over the SDS and SSS and the regional MFR values. Cardiac entropy has been studied in the past, and Costa et al. concluded that heart rate complexity decreased with the increase

of age, but also with disease[64]. A review by Chen et al. summarizes a number of articles about entropy for coronary heart disease. The consensus from five articles was that patients with CAD had lower entropy values compared to healthy subjects. These studies measured various type of entropy, including entropy of diastolic heart sounds, entropy of the heart rate and cardiac magnetic field mapping. If the level of entropy is also predictive for the severity of CAD, it can be explained that MPE is a predictor for obstructive CAD.

Medication usage in general was not ranked as a strong predictor for an obstructive event. The pharmacological management of CAD is aimed to reduce anginal symptoms and to prevent cardiovascular events, and since it is the primary treatment for patients with low risk, medication usage alone is probably not an indicator for the occurrence of cardiovascular events since both patients who are more likely as well as patients who are less likely to suffer a cardiovascular event are prescribed the same medication. It may be possible that differences in medication dosages are able to distinguish high risk patients from low risk patients, but this information was not accessible to us.

Not just medication categories were ranked as weak predictors. The type of gender and non-cardiac medical history did not show any relationship with obstructive CAD. This is surprising since the contrary was concluded by Reeh at al., who found that male gender had an odds ratio of 4.25 (95% CI 3.13–5.77) for the prediction of obstructive CAD[41]. In fact, male gender is a known risk factor for obstructive CAD, and has been used extensively in PTP-risk models, such as the established Diamond-Forrester risk score and the Duke criteria for chest pain[65, 66]. Our feature importance ranking suggests that the model would perform equally well when features like gender and non-cardiac medical history are not used during training.

No direct comparison can be made with previous ML approaches for the prediction of obstructive CAD because either the variety of input features or the specific prediction task differed. In a recent study by Hu et al, ML was used to predict per-vessel early coronary revascularization after SPECT MPI, a LogitBoost model was implemented and trained with 10-fold crossvalidation[42]. In this study, a combination of MPI derived parameters and clinical features was used. A notable difference is that in this study, SPECT MPI was used and not PET PET. Similarities with regards to predictors can be seen: Similar to what we found, MPI derived parameters were superior predictors compared to the classical clinical risk factors and other clinical parameters. The best performing models reached an AUC of 0.79. We were able to obtain an AUC of 0.90. The main difference between this study and our study is that we chose to predict long term outcome, and not just early revascularization. Also, our AUC value was by large determined by the specificity, and since the dataset that we used to train contained a strong imbalance towards negative cases, a higher AUC is obtained more easily. Only part of our obstructive events can be considered as early revascularization.

Juarez-Orozco et al. implemented ML algorithms to predict myocardial ischemia and the risk of major adversarial cardiac events (MACE). However, here the PET derived quantification was used to classify patient outcome, and not used as input for the model. A combination of patient characteristics and cardiac parameters were used to model for the outcome. In their study LogitBoost was found to be the best performing algorithm out of a number of candidate algorithms, including Naïve Bayes, Random Forest and Logistic Regression[37].

Haro Alonso et al. developed ML models to predict cardiac death on a dataset of 8321 patients. Opposed to the previous study, this time the input features included SPECT MPI findings. Their best performing algorithm was a SVM which reached an AUC of 0.83[40].

4.4.1 Strenghts and limitations

The integration of MPI derived features with CAC scoring and clinical variables is what distinguishes this study from previously performed studies. The incremental prognostic information of a CAC scoring, or another form of characterizing the atherosclerotic burden should always be incorporated if this kind of risk modeling is available. Note that the obvious disadvantage of CAC scoring is that it does not exclude risk caused by soft plaques.

This study has several limitations. First of all, arguably we could have improved our model even further by adding ECG derived input features such as presence of atrial fibrillation or left bundle branch block. Other features that may be included are PTP and anginal symptoms. However, our work resulted in perhaps the most detailed risk model to date, and never before has MPE, as derived from regional CFR measurements, been used in a predictive model.

All cause death was not considered as an obstructive CAD event. Death as a result of myocardial infarction should be classified as an obstructive event. However, we found that the cause of death was rarely well documented in the patient records that were available. And a well-documented cause of death is essential in our case since cardiac death can be attributed to more causes than just myocardial infarction. Despite the large study population, the number of obstructive events was relatively low at 11%. This meant that we had to work with a strongly imbalanced dataset, and performance metrics such as accuracy will not give proper representation of the actual performance. Unfortunately this caused the confidence intervals of especially the sensitivity to become quite large. The primary strength of the current XGBoost model lies in the specificity, as it can be used to exclude patients from further examinations.

In our grid search we limited ourselves to a finite number of hyperparameter values. The benefit of performing a grid search is that good results can be obtained in a short time span. However, the final hyperparameters may not be the optimal hyperparameters. A framework that is able to iteratively optimize the hyperparameters based on the results of the previous result, and actively searches for the best parameters may be advantageous[67, 68]. The obvious downside is the additional required computational time whilst the benefit is not guaranteed.

All results reported in this work originated from models that were trained using 5-fold cross-validation. The performance is validated on a dataset that is specific to our hospital. A dataset from other medical centres are required to evaluate how well the model generalizes beyond data from Isala.

4.4.2 Clinical implementation

The XGboost is able to provide a prediction of the occurrence of an obstructive event after MPI. Therefore, it may be utilized to assist clinicians as a risk stratification tool. It may be able to offer more assurance before the decision of further evaluation by means of invasive CAG is made. According to the performance results, this model is able to provide similar performance compared to the current clinical standard. The performance of the model is especially accurate in cases which are regarded by the model as low risk. The integration of this model into the diagnostic workflow can therefore be beneficial. First of all, the tool can be consulted without much effort as a second opinion, potentially improving the overall diagnostic accuracy and preventing needless ICA procedures. The model can accurately identify or verify low risk patients and prevent redundant follow-up. Secondly, the model can be used as a teaching tool. Guidance provided by the model to unexperienced clinicians may speed up the learning process and allow them to reach expert levels sooner.

Furthermore, it may be interesting to investigate if models with similar performance can be developed with less features. A surplus of irrelevant features can cause models to overfit more easily[69]. Moreover, potential simplification of these models is certainly interesting in the context of clinical implementation. As a starting point, the feature importance ranking can be used to determine a cut-off for features.

4.5 Conclusion

We have implemented multiple ML algorithms to develop a model for the prediction of obstructive CAD, based on a combination of clinical and quantitative features. According to our results XGboost outperformed other algorithmic approaches, and performed equally well as a clinician. The resulting model is able to provide individual risk stratification by predicting the possibility of an obstructive cardiovascular event. This indicates the potential of using this model as a decision support tool to. The model can be implemented to reduce the number of patients sent for ICA. However, before that is a possibility, further validation of the model and seamless integration within the diagnostic workflow are required.

Chapter 5

Future Perspectives

For the past year I worked on a clinically relevant problem that I attempted to tackle with a ML approach. Though the results seem promising, further improvements can be addressed considering the development and clinical implementation.

First of all, and this applies to all projects in ML, there can never be enough data. The amount of data used in this study was acceptable for scientific/medical research purposes, however for ML, the dataset was quite small. Also, the dataset was heavily imbalanced towards non-obstructive CAD cases. This is not necessarily problematic when developing models. However, the single best step to improve predictive models is developing them with a larger dataset. Moreover, since we are primarily interested in detecting obstructive CAD, ideally the dataset should include more of these training examples. There are methods to (synthetically) increase the volume of a dataset in this specific task.

First of all, consortia can be formed of multiple medical centers that pool their resources together. There are opportunities in cooperation regarding the development of models, where sometimes expensive hardware is required. Improvements can be made regarding the sharing of medical data. Even though data sharing between institutions is becoming more common, it is still a notoriously difficult process. The important but strict national and European regulations regarding patient privacy make the sharing of data more complicated compared to other industries, but also compared to other regions in the world. There have been some initiatives to share medical data in a responsible manner. On a European scale, some of the most noteworthy being the European Medical Information Framework, the European Health Research and Innovation Cloud and the European Health Data and Evidence Network[70, 71]. Databases are also being established on a global scale. Although most of these databases (e.g. Cancer Imaging Archive, the Osteoarthritis Initiative and the National Institute of Arthritis and Musculoskeletal and Skin Diseases) focus on one specific item, it is a step towards more thorough collaboration[72].

The second approach to gain more data for the development of models is synthetic data generation. The generation of new data can be seen as some form of data augmentation. Data augmentation is a technique used to reduce overfitting in ML, but it can also be used to generate synthetic data entries that satisfy statistical properties of the original dataset. The newly generated data entries can be combined with the original ones to enhance the variability within the dataset[72]. Another technique uses synthetic data for heavily imbalanced datasets: the ADASYN algorithm. This algorithm, as proposed by He et al., aims to improve the learning by reducing the bias as a result of the class imbalance and focusing on difficult examples. This is achieved by the generation of data samples that lie next to original observations

which are wrongly classified by an algorithm[73]. It must be said that the usage of generated data in healthcare should be handled with care since unwanted biases can be introduced easily. Still, both techniques can be applied on the dataset used in this thesis.

Another approach to combat the lack of obstructive CAD cases is to modify the inclusion criteria. We opted to exclude patients with a cardiovascular history since the perfusion quantification of patients that underwent PCI or CABG is not verified. However, including this patient population can be interesting for two reasons. First of all the size of the dataset will increase. More importantly, in theory, a more versatile model can be developed.

Alongside the prediction of obstructive CAD for patients with prior PCI's and CABG's a number of other improvements can be made to this model in the future. First of all, there lies a challenge in the prediction of regional obstruction. However, per-vessel prediction requires per-vessel labeling. In this thesis, a combination of (semi-)quantitative and qualitative features was used. Since the quantitative features are extracted from a scan, there is an opportunity in using the raw data of the PET scan as additional input for the model. There is a possibility that predictive information can be extracted from the raw data by well-established techniques from the domain of deep learning such as a convolutional neural network(CNN)[74]. The raw data can be processed by a 2-dimensional or 3-dimensional CNN and the outcome(s) can be integrated within the current XGBoost model as additional input.

Clinical implementation occurs only after extensive validation. However, the development of ML models is only one aspect of implementation. These AI based models to be used as tools for healthcare should merge flawlessly into the current workflow. The majority of tools developed for healthcare fail in this aspect, and simply remain underused or unused after implementation. Sometimes there is a sense of reluctance to adapt. To overcome this, the tool should be user friendly, fast and not require a lot of effort to use. User friendliness can be achieved by intuitive design choices and integration in existing frameworks such as the electronic health record system(EHRS) or the radiology information system (RIS). More ideally, the tool can provide fully automatic risk evaluation. As of now this is impossible since it requires the coupling of multiple self-contained programs such as the EHRS and the RIS. This is not yet possible but it is not unimaginable in the near future.

I believe that the technical physician has an important role in all aspects within this field. This includes the development of models, the validation of the models and the clinical implementation. It is essential to have knowledge about a specific domain for which an algorithm is developed. The combination of knowledge that a technical physician has about a deep understanding of anatomy, pathology, healthcare organization and patient workflow is in that sense unique.

To conclude, it is important to realize that the amount of available data will never decrease, and the availability of computational power will increase. ML can be utilized to advance healthcare as never before. ML can address challenges by improving quality of care, reducing healthcare costs, supporting workload management and reducing human error, while providing an unparalleled personalized approach. Difficulties remain, but are being addressed on a worldwide scale. The potentiality for improvement is huge for ML in healthcare.

Appendix A

Description of the Kaggle-dataset

TABLE A.1: Features of the Kaggle-dataset

Feature	**Possible values [range]**
Male	[0, 1]
age	[15...23]
RG(00,0,0.7),weekly study time	[1,2,3,4]
RG(00,0,0.7),number of past class failures	[0,1,2,3]
extra educational support	[0, 1]
RG(00,0,0.7),family educational support	[0, 1]
extra-curricular activities	[0, 1]
strives for higher education	[0, 1]
currently in a romantic relationship	[0, 1]
quality of family relationships	[0,1,2,3,4,5]
amount of free time	[0,1,2,3,4,5]
goes out with friends	[0,1,2,3,4,5]
Daily alcohol consumption	[0,1,2,3,4,5]
Weekly alcohol consumption	[0,1,2,3,4,5]
Health status	[0,1,2,3,4,5]
Number of absences	[0 ... 75]
Math grade term 1	[1 ... 20]
Math grade term 2	[1 ... 20]
Math grade term 3	[1 ... 20]
Travel time to school	[1,2,3,4]
Paid educational support	[0, 1]
activities	[0, 1]
nursery	[0, 1]
internet	[0, 1]

Appendix B

Abstract for EANM2020

Topic: B4 CARDIOVASCULAR IMAGING CLINICAL STUDY B41 PERFUSION
Title: Identification of predictors for myocardial blood flow using Rubidium-82 PET
This abstract was accepted for oral presentation for the EANM2020

Authors: RJ Metselaar[1,2], JA van Dalen[3], BN Vendel[2], JR Timmer[4], M Mouden[4], JD van Dijk[2]
[1]Technical Medicine, University of Twente
[2]Isala Ziekenhuis, Department of Nuclear Medicine
[3]Isala Ziekenhuis, Department of Medical Physics
[4]Isala Ziekenhuis, Department of Cardiology

Introduction Myocardial blood flow (MBF) measurements using Rubidium-82 PET provides incremental diagnostic and prognostic information in the evaluation of coronary artery disease (CAD). In particular, high myocardial flow reserve (MFR) values have a high negative predictive value for obstructive CAD. Understanding which features are characteristic for normal MBF and MFR values can aid in the risk stratification of patients with suspected obstructive CAD. The aim of this study was to identify these features for developing a clinical decision tool.

Materials and Methods We retrospectively included 997 patients with suspected CAD (50,7% Male, age 65,99 ± 10.43 years), who were referred for rest and regadenoson-induced stress Rubidium-82 PET/CT (Discovery 690, GE Healthcare). Cardiac risk factors; cigarette smoking, hypertension, hypercholesterolemia, diabetes, positive family history of CAD; prior medical history; age; gender; body mass index (BMI); creatinine serum values; coronary artery calcification (CAC) score and medication usage were registered at time of the PET/CT examination. All variables except CAC score, age, resting heart rate and BMI were transformed to categorical variables. Multiple linear regression with forward stepwise selection (F-to-enter < 0.050) was calculated to correlate variables with rest and stress MBF and MFR. The importance of significant predictors was determined by ranking using the incremental increase in adjusted R^2.

Results Significant predictors for high rest MBF were, in order of importance, high heart rate, female sex, calcium-channel blocker usage, low BMI, calcium-channel blocker usage, hypertension, no acenocoumerol usage and older age. The adjusted R2 was 0.34. Furthermore, the significant predictors for high stress MBF were, in order of importance, female sex, high heart rate, low CAC score, low BMI, younger age, clopidogrel usage, no beta blocker usage, COPD. (Adjusted R2: 0.23). The predictors for high MFR were younger age, low heart rate, low CAC score, no calcium-channel blocker usage, no underlying diabetes mellitus, positive family history of CAD normal creatinine serum levels and male. (Adjusted R2: 0.19).

Conclusion We identified several features that correlate with MBF and MFR measurements using Rubidium-82 PET. These features can be used in a clinical decision tool for risk stratification of patients with CAD. Further studies are needed to determine the prognostic value of the identified features in combination with MBF and MFR.

Bibliography

(1) Mendis, S.; Puska, P.; Norrving, B.; Organization, W. H., et al., *Global atlas on cardiovascular disease prevention and control*; World Health Organization: 2011.

(2) Van Dis S.J.and Deckers, J.; Poos M.J.J.C. andNielen, M. a. P.; Gommer, A.; Rodriguez, M. Coronaire Hartziekten in Nederland, year = 2020, url = https://www.volksgezondheiden hartziekten/cijfers-context/huidige-situatie, urldate = 2020-10-23.

(3) Wang, H; Naghavi, M; Allen, C; Barber, R.; Carter, A; Casey, D.; Charlson, F.; Chen, A.; Coates, M.; Coggeshall, M, et al. *The Lancet* **2016**, *388*, 1459–1544.

(4) Alexopoulos, N.; Raggi, P. *Nature Reviews Cardiology* **2009**, *6*, 681–688.

(5) Schoenhagen, P.; Ziada, K. M.; Kapadia, S. R.; Crowe, T. D.; Nissen, S. E.; Tuzcu, E. M. *Circulation* **2000**, *101*, 598–603.

(6) Fiss, D. M. *Applied radiology* **2007**, 14.

(7) On Myocardial Segmentation, A. H. A. W. G.; for Cardiac Imaging: R.; Cerqueira, M. D.; Weissman, N. J.; Dilsizian, V.; Jacobs, A. K.; Kaul, S.; Laskey, W. K.; Pennell, D. J.; Rumberger, J. A.; Ryan, T., et al. *Circulation* **2002**, *105*, 539–542.

(8) Dilsizian, V.; Bacharach, S. L.; Beanlands, R. S.; Bergmann, S. R.; Delbeke, D.; Dorbala, S.; Gropler, R. J.; Knuuti, J.; Schelbert, H. R.; Travin, M. I. *Journal of Nuclear Cardiology* **2016**, *23*, 1187–1226.

(9) Knuuti, J.; Wijns, W.; Saraste, A.; Capodanno, D.; Barbato, E.; Funck-Brentano, C.; Prescott, E.; Storey, R. F.; Deaton, C.; Cuisset, T., et al. *European heart journal* **2020**, *41*, 407–477.

(10) Demir, O. M.; Alfakih, K.; Plein, S. Current international guidelines for the investigation of patients with suspected coronary artery disease, 2014.

(11) Neves, P. O.; Andrade, J.; Monção, H. *Radiologia brasileira* **2017**, *50*, 182–189.

(12) Agatston, A. S.; Janowitz, W. R.; Hildner, F. J.; Zusmer, N. R.; Viamonte, M.; Detrano, R. *Journal of the American College of Cardiology* **1990**, *15*, 827–832.

(13) Sarwar, A.; Shaw, L. J.; Shapiro, M. D.; Blankstein, R.; Hoffman, U.; Cury, R. C.; Abbara, S.; Brady, T. J.; Budoff, M. J.; Blumenthal, R. S., et al. *JACC: Cardiovascular Imaging* **2009**, *2*, 675–688.

(14) Williams, M. C.; Moss, A. J.; Dweck, M.; Adamson, P. D.; Alam, S.; Hunter, A.; Shah, A. S.; Pawade, T.; Weir-McCall, J. R.; Roditi, G., et al. *Journal of the American College of Cardiology* **2019**, *73*, 291–301.

(15) Skinner, J. S.; Smeeth, L.; Kendall, J. M.; Adams, P. C.; Timmis, A. *Heart* **2010**, *96*, 974–978.

(16) Kalisz, K.; Buethe, J.; Saboo, S. S.; Abbara, S.; Halliburton, S.; Rajiah, P. *Radiographics* **2016**, *36*, 2064–2083.

(17) Rizvi, A.; Han, D.; Danad, I.; Hartaigh, B. Ó.; Lee, J. H.; Gransar, H.; Stuijfzand, W. J.; Roudsari, H. M.; Park, M. W.; Szymonifka, J., et al. *JACC: Cardiovascular Imaging* **2018**, *11*, 589–599.

(18) Hulten, E. A.; Carbonaro, S.; Petrillo, S. P.; Mitchell, J. D.; Villines, T. C. *Journal of the American College of Cardiology* **2011**, *57*, 1237–1247.

(19) Delcour, K. S.; Khaja, A.; Chockalingam, A.; Kuppuswamy, S.; Dresser, T. *Angiology* **2009**, *60*, 318–321.

(20) Shaw, L. J.; Iskandrian, A. E. *Journal of nuclear cardiology* **2004**, *11*, 171–185.

(21) Schindler, T. H.; Schelbert, H. R.; Quercioli, A.; Dilsizian, V. *JACC: Cardiovascular Imaging* **2010**, *3*, 623–640.

(22) Nakazato, R.; Berman, D. S.; Alexanderson, E.; Slomka, P. *Imaging in medicine* **2013**, *5*, 35.

(23) Danad, I.; Raijmakers, P. G.; Driessen, R. S.; Leipsic, J.; Raju, R.; Naoum, C.; Knuuti, J.; Mäki, M.; Underwood, R. S.; Min, J. K., et al. *JAMA cardiology* **2017**, *2*, 1100–1107.

(24) Ghotbi, A. A.; Kjær, A.; Hasbak, P. *Clinical physiology and functional imaging* **2014**, *34*, 163–170.

(25) Herrero, P.; Markham, J.; Shelton, M. E.; Weinheimer, C. J.; Bergmann, S. R. *Circulation* **1990**, *82*, 1377–1386.

(26) Gould, K. L. In *Seminars in nuclear medicine*, 1987; Vol. 17, pp 121–130.

(27) Al Jaroudi, W.; Iskandrian, A. E. *Journal of the American College of Cardiology* **2009**, *54*, 1123–1130.

(28) Yoshida, K.; Mullani, N.; Gould, K. L. *Journal of Nuclear Medicine* **1996**, *37*, 1701–1712.

(29) Herrero, P.; Markham, J.; Shelton, M. E.; Bergmann, S. R. *Circulation research* **1992**, *70*, 496–507.

(30) Huang, S.-C.; Williams, B.; Krivokapich, J; Araujo, L; Phelps, M. E.; Schelbert, H. R. *American Journal of Physiology-Heart and Circulatory Physiology* **1989**, *256*, H1156–H1164.

(31) Lortie, M.; Beanlands, R. S.; Yoshinaga, K.; Klein, R.; DaSilva, J. N.; DeKemp, R. A. *European journal of nuclear medicine and molecular imaging* **2007**, *34*, 1765–1774.

(32) Murthy, V. L.; Bateman, T. M.; Beanlands, R. S.; Berman, D. S.; Borges-Neto, S.; Chareonthaitawee, P.; Cerqueira, M. D.; DeKemp, R. A.; DePuey, E. G.; Dilsizian, V., et al. *Journal of Nuclear Cardiology* **2018**, *25*, 269–297.

(33) Naya, M.; Murthy, V. L.; Taqueti, V. R.; Foster, C. R.; Klein, J.; Garber, M.; Dorbala, S.; Hainer, J.; Blankstein, R.; Resnic, F., et al. *Journal of Nuclear Medicine* **2014**, *55*, 248–255.

(34) Hsiao, E.; Ali, B.; Blankstein, R.; Skali, H.; Ali, T.; Bruyere Jr, J.; Kwong, R. Y.; Di Carli, M. F.; Dorbala, S. *Journal of nuclear medicine: official publication, Society of Nuclear Medicine* **2013**, *54*, 1748.

(35) Zampella, E.; Acampa, W.; Assante, R.; Nappi, C.; Gaudieri, V.; Mainolfi, C. G.; Green, R.; Cantoni, V.; Panico, M.; Klain, M., et al. *European journal of nuclear medicine and molecular imaging* **2018**, *45*, 521–529.

(36) Arsanjani, R.; Xu, Y.; Dey, D.; Vahistha, V.; Shalev, A.; Nakanishi, R.; Hayes, S.; Fish, M.; Berman, D.; Germano, G., et al. *Journal of Nuclear Cardiology* **2013**, *20*, 553–562.

(37) Juarez-Orozco, L. E.; Knol, R. J.; Sanchez-Catasus, C. A.; Martinez-Manzanera, O.; Van der Zant, F. M.; Knuuti, J. *Journal of Nuclear Cardiology* **2020**, *27*, 147–155.

(38) Simeone, O. *arXiv preprint arXiv:1709.02840* **2017**.

(39) Friedman, J.; Hastie, T.; Tibshirani, R., *The elements of statistical learning*; 10; Springer series in statistics New York: 2001; Vol. 1.

(40) Alonso, D. H.; Wernick, M. N.; Yang, Y.; Germano, G.; Berman, D. S.; Slomka, P. *Journal of Nuclear Cardiology* **2019**, *26*, 1746–1754.

(41) Reeh, J.; Therming, C. B.; Heitmann, M.; Højberg, S.; Sørum, C.; Bech, J.; Husum, D.; Dominguez, H.; Sehestedt, T.; Hermann, T., et al. *European Heart Journal* **2019**, *40*, 1426–1435.

(42) Hu, L.-H.; Betancur, J.; Sharir, T.; Einstein, A. J.; Bokhari, S.; Fish, M. B.; Ruddy, T. D.; Kaufmann, P. A.; Sinusas, A. J.; Miller, E. J., et al. *European Heart Journal-Cardiovascular Imaging* **2020**, *21*, 549–559.

(43) Al'Aref, S. J.; Maliakal, G.; Singh, G.; van Rosendael, A. R.; Ma, X.; Xu, Z.; Alawamlh, O. A. H.; Lee, B.; Pandey, M.; Achenbach, S., et al. *European Heart Journal* **2020**, *41*, 359–367.

(44) Motwani, M.; Dey, D.; Berman, D. S.; Germano, G.; Achenbach, S.; Al-Mallah, M. H.; Andreini, D.; Budoff, M. J.; Cademartiri, F.; Callister, T. Q., et al. *European heart journal* **2017**, *38*, 500–507.

(45) Cortez, P.; Silva, A. M. G. **2008**.

(46) Tibshirani, R. *Journal of the Royal Statistical Society: Series B (Methodological)* **1996**, *58*, 267–288.

(47) Cortes, C.; Vapnik, V. *Machine learning* **1995**, *20*, 273–297.

(48) Rokach, L.; Maimon, O. Z., *Data mining with decision trees: theory and applications*; World scientific: 2008; Vol. 69.

(49) Friedman, J. H. *Annals of statistics* **2001**, 1189–1232.

(50) Zhang, C.; Liu, C.; Zhang, X.; Almpanidis, G. *Expert Systems with Applications* **2017**, *82*, 128–150.

(51) Organization, W. H. et al. *World Health Organization* **2019**.

(52) Townsend, N.; Wilson, L.; Bhatnagar, P.; Wickramasinghe, K.; Rayner, M.; Nichols, M. *European heart journal* **2016**, *37*, 3232–3245.

(53) Ziadi, M. C.; deKemp, R. A.; Williams, K.; Guo, A.; Renaud, J. M.; Chow, B. J.; Klein, R.; Ruddy, T. D.; Aung, M.; Garrard, L., et al. *Journal of Nuclear Cardiology* **2012**, *19*, 670–680.

(54) Djaileb, L; Legagneur, C; De Leiris, N; Leenhardt, J; Desvigne, M; Fagret, D; Riou, L; Rochette, G. B. *Médecine Nucléaire* **2020**, *44*, 102.

(55) Chen, C.; Jin, Y.; Lo, I. L.; Zhao, H.; Sun, B.; Zhao, Q.; Zheng, J.; Zhang, X. D. *International journal of biological sciences* **2017**, *13*, 1320.

(56) Okuda, K.; Nakajima, K. What does entropy reveal in phase analysis of myocardial perfusion SPECT?, 2019.

(57) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V., et al. *the Journal of machine Learning research* **2011**, *12*, 2825–2830.

(58) Mann, H. B.; Whitney, D. R. *The annals of mathematical statistics* **1947**, 50–60.

(59) Glass, G. V. *American Educational Research Journal* **1966**, *3*, 187–190.

(60) Vollset, S. E. *Statistics in medicine* **1993**, *12*, 809–824.

(61) Chang, S. M.; Nabi, F.; Xu, J.; Peterson, L. E.; Achari, A.; Pratt, C. M.; Mahmarian, J. J. *Journal of the American College of Cardiology* **2009**, *54*, 1872–1882.

(62) Rozanski, A.; Slomka, P.; S. Berman, D. Extending the use of coronary calcium scanning to clinical rather than just screening populations: ready for prime time?, 2016.

(63) Catapano, A. L.; Graham, I.; De Backer, G.; Wiklund, O.; Chapman, M. J.; Drexel, H.; Hoes, A. W.; Jennings, C. S.; Landmesser, U.; Pedersen, T. R., et al. *European heart journal* **2016**, *37*, 2999–3058.

(64) Costa, M. D.; Goldberger, A. L. *Entropy* **2015**, *17*, 1197–1203.

(65) Diamond, G. A.; Forrester, J. S. *New England Journal of Medicine* **1979**, *300*, 1350–1358.

(66) Pryor, D. B.; Harrell Jr, F. E.; Lee, K. L.; Califf, R. M.; Rosati, R. A. *The American journal of medicine* **1983**, *75*, 771–780.

(67) Bergstra, J.; Bengio, Y. *The Journal of Machine Learning Research* **2012**, *13*, 281–305.

(68) Bergstra, J. S.; Bardenet, R.; Bengio, Y.; Kégl, B. In *Advances in neural information processing systems*, 2011, pp 2546–2554.

(69) Raymer, M. L.; Punch, W. F.; Goodman, E. D.; Kuhn, L. A.; Jain, A. K. *IEEE transactions on evolutionary computation* **2000**, *4*, 164–171.

(70) Lovestone, S.; Consortium, E. *Learning Health Systems* **2020**, *4*, e10214.

(71) Aarestrup, F. M.; Albeyatti, A.; Armitage, W.; Auffray, C; Augello, L.; Balling, R.; Benhabiles, N.; Bertolini, G.; Bjaalie, J.; Black, M., et al. *Genome medicine* **2020**, *12*, 1–14.

(72) Ranschaert, E. R.; Morozov, S.; Algra, P. R., *Artificial Intelligence in Medical Imaging: Opportunities, Applications and Risks*; Springer: 2019.

(73) He, H.; Bai, Y.; Garcia, E. A.; Li, S. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, 2008, pp 1322–1328.

(74) Krizhevsky, A.; Sutskever, I.; Hinton, G. E. *Communications of the ACM* **2017**, *60*, 84–90.