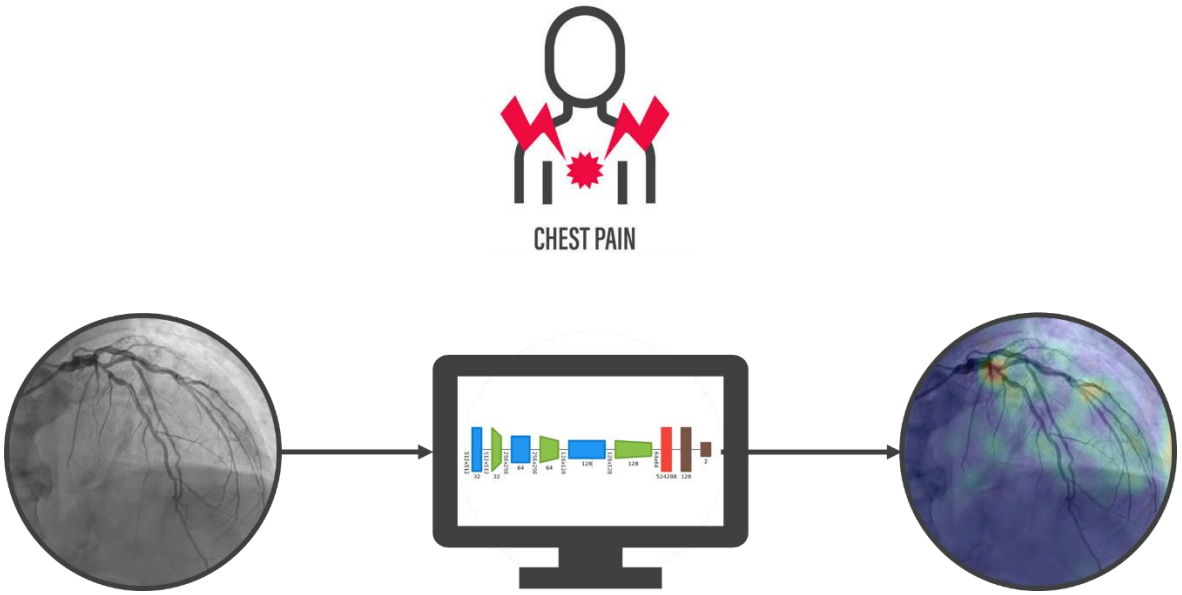
The future of interventional cardiology: Machine learning algorithms for solving diagnostic and therapeutic challenges



W. P. M. van der Loo
Technical Medicine | Imaging & Intervention
September 2020 – June 2021

MASTER'S THESIS

The future of interventional cardiology: Machine learning algorithms for solving diagnostic and therapeutic challenges

by

W.P.M. Van der Loo

Student number 4375394

Final version | June 16, 2021

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University | Delft University of Technology | Erasmus University Rotterdam

Master thesis project (TM30004; 35 ECTS)

December 2020 – June 2021

Supervisor(s):

Medical supervisor: Dr. R.W.C. Scherptong

Department of Cardiology

Technical supervisor: Dr. Ir. J. Dijkstra

Division of image processing

Thesis committee members:

Prof. dr. ir. Jaap Harlaar, TU Delft (chair)

Dr. R. W. C. Scherptong, LUMC

Dr. Ir. J. Dijkstra, LUMC

"Without data you are just another person with an opinion"

W. Edwards Deming (1900 – 1993)
Statistician

Preface

Almost 7 years ago, I started the new Bachelor's program Clinical Technology in Delft, Leiden and Rotterdam. Starting a new Bachelor's program was both exciting and hard at the same time. However, it was not a hard choice for me to start this journey. Combining medicine with technological innovations in a clinical context sounded about as good as it gets to me. During my Bachelor I found out that I was especially interested in image processing and image guided interventions. Fortunately, there was a master track specially build around imaging and intervention.

In the second year of my master's program, I did several internships in various clinical departments. During these internships I found out that I also really enjoyed the medical parts of the internship. After thinking about this a lot and talking to several technical physicians, I decided to combine my Technical Medicine study with the Medicine program. After finishing the medicine pre-master last year, I started my master thesis project at the department of cardiology in the LUMC. A department that I find very interesting, not only because I am fascinated by the whole cardiovascular system but also because high-end technology is combined with healthcare in a very diverse way.

During my project, I focused on identifying how machine learning can improve the clinical practice of interventional cardiology. Both subjects that I find really interesting. This project also really showed the added value of technical physicians to me: bringing the medical and technical worlds together by being able to communicate with professionals of both sides on their level. In my opinion, this is essential when translating highly technical innovations such as machine learning to the clinical practice. At first, I found it hard to see this and I want to thank Roderick Scherptong for his efforts in making this clear to me. Furthermore, I want to thank my supervisors Roderick Scherptong and Jouke Dijkstra for their extensive feedback and support during the whole project. I also would like to thank all the respondents I interviewed for my thesis feasibility study, the people at LKEB who helped me with setting up a workstation and an environment for creating a deep learning network, the software engineering team at the cardiology department for helping me with querying data from EPD-Vision and everybody at the department for their enthusiasm and support during the project.

I am looking forward to two more years of clinical internships and hopefully after this I can combine my passion for technology and medicine in clinical practice.

Wouter van der Loo
June 16, 2021
The Hague

Table of contents

List of abbreviations

AI: Artificial intelligence
ANN: Artificial neural network
AP: Angina pectoris

CABG: Coronary artery bypass grafting
CAD: Coronary artery disease
CNN: Convolutional neural network
CVD: Cardiovascular disease

DL: Deep learning
DICOM: Digital Imaging and Communications in Medicine

EHR: Electronic health records

FCN: Fully convolutional neural network
FFR: Fractional flow reserve

GAN: Generative adversarial network

ICA: Invasive coronary angiography
iFR: Instantaneous wave-free ratio
INOCA: Ischemia with non-obstructed coronary arteries

LAD: Left anterior descending coronary artery
LSTM-RNN: Long short-term memory recurrent neural network

MINOCA: Myocardial infraction with non-obstructed coronary arteries
ML: Machine learning

NLP: Natural language processing
NSTEMI: Non-ST elevation myocardial infarction

PTCA: Percutaneous transluminal coronary angioplasty
PCI: percutaneous coronary intervention

QCA: Quantitative coronary angiography
RCA: Right coronary artery
RCx: Ramus circumflex

SCAD: Spontaneous coronary artery dissection
SQL: Structured query language
STEMI: ST elevation myocardial infarction
SVM: Support vector machine
RNN: Recurrent neural network

1. General introduction

1.1 Medical background

1.1.1 Coronary artery disease

Cardiovascular disease (CVD) is a group of diseases that affect both the heart and blood vessels. (1) This includes coronary artery disease (CAD). In CAD, the coronary arteries, that supply the myocardium with oxygen, are narrowed or even blocked by an inflammatory process called atherosclerosis. (2) A schematic representation is shown in figure 1.1. This process develops from inflammation in the blood vessel wall with accumulation of lipids and fibrous elements. Atherosclerosis has a complex etiology, however, numerous risk factors have been identified. These risk factors can generally be grouped into genetic and environmental components. Important factors with a strong genetic component are hypercholesterolemia, hypertension, family history, diabetes and obesity. Important environmental factors are a high-fat diet, smoking, low levels of antioxidants, aging and lack of exercise. (3) In CAD, symptoms may arise when there is a reduced blood flow, and so a reduced oxygen supply, to the myocardium. These symptoms are chest pain, often related to exercise, pain in the arms or shoulders and dyspnea.

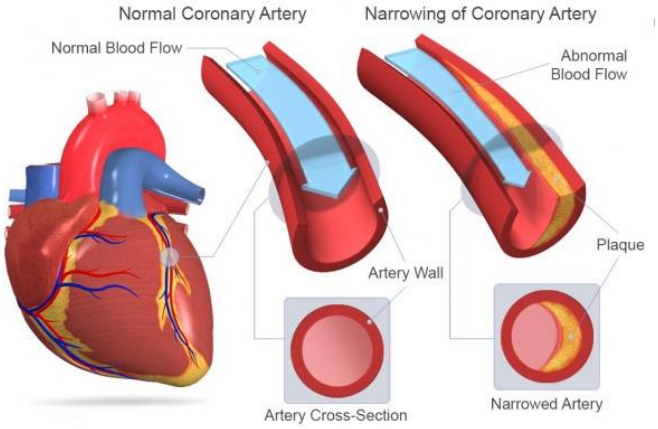


Fig. 1.1 Coronary artery disease, a schematic representation. (4)

CVD is one of the leading causes of death and disability worldwide and the incidence is still increasing. (5) Mortality has been gradually decreasing over the last decades in Western countries, due to better prevention by reducing risk factors and improvements in medical therapy and interventional options. (6) In 2019 in the Netherlands there were an estimate of 773.000 people with CAD of which 62% was male. (7) CAD also was responsible for the highest burden of disease in Netherlands for 2019. (7)

1.1.2 Coronary anatomy

The coronary arteries can generally be divided in a left and a right coronary artery. The left originates from the left sinus Valsalva and starts with the left main coronary artery (LM), the LM bifurcates into the circumflex coronary artery (RCx; ramus circumflex) and the left anterior descending artery (LAD). In some cases, a third artery is present at this point: the intermediate or anterolateral artery, making the bifurcation a trifurcation. The LAD is positioned on the anterior interventricular sulcus with small branches penetrating the septum. The RCx is positioned in the left atrioventricular sulcus and it has one or more obtuse marginal branches (OM). The right coronary artery (RCA) originates from the right sinus Valsalva. The RCA is positioned in the right atrioventricular sulcus and can have acute marginal branches. (8) A schematic overview of the coronary anatomy is shown in figure 1.2. The artery that supplies the posterior descending artery determines the dominance. Approximately 70% of the general population is right dominant (posterior descending artery from the RCA), 10% is left dominant (branch from the RCx) and 20% is co-dominant, which means that it is supplied by both the RCA and the RCx.

(9) There are several relatively common anatomical variations possible. Such as the presence of an intermediate branch. It is also possible that the LAD and the RCx have different ostia and that thus no LM is present. This is detected in 0.4-8% of the general population. (10)

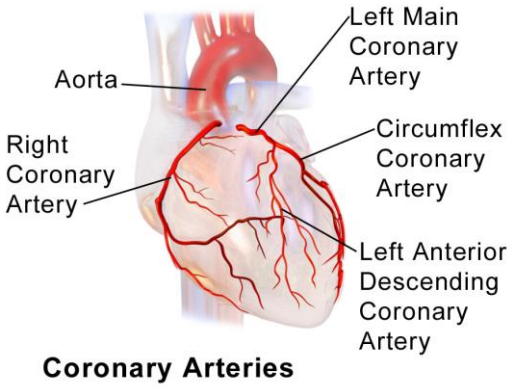


Fig. 1.2 Anatomy of the coronary arteries. (4)

1.1.3 Diagnostic process for CAD

The diagnostic work-up for patients suspected of CAD consists of selecting testing modalities based on the likelihood of disease. It starts with the assessment of symptoms and clinical investigation. This is an important step in the process of diagnosis, as it is possible to achieve a high degree of certainty based on history alone. Physical examination and diagnostic testing are often needed to conform the diagnosis and assesses the severity of disease. Before starting any form of diagnostic testing, the patient's general health, comorbidities and quality of life should be considered. If a patient is unlikely to be eligible for revascularization, this can reduce the need for further testing to a minimum. The next step is basic testing. This consists of biochemical testing, a resting ECG, possibly ambulatory ECG monitoring, resting echocardiography and a chest X-ray in some patients. Cardiac magnetic resonance (CMR) imaging may also be considered when the echocardiogram is inconclusive. (11) Before any further testing, the pre-test probability and clinical likelihood of CAD should be determined. Diagnostic testing is most useful in patients with an intermediate likelihood. A predictive model can be used to determine pre-test probability based on age, sex and nature of symptoms. Incorporation of risk factors for CVD, resting ECG changes and coronary calcification increases the accuracy of predicting the presence of CAD. (12)

Non-invasive testing

Next the appropriate test should be selected. This can be done on clinical likelihood and the condition of the patient. For patients that have no benefit from revascularization, due to comorbidities and overall quality of life, clinical diagnosis of CAD combined with medical therapy is sufficient. Patients with a high likelihood of CAD, unresponsiveness to medical therapy or typical complaints at a low level of exercise combined with a high event risk based on initial evaluation can directly be referred for invasive coronary angiography (ICA). Hemodynamic significance of stenoses can invasively be confirmed. (13, 14) In patients with a higher degree of uncertainty, non-invasive assessment can be used for diagnosis and determining event risk. This can be done by non-invasive functional testing or anatomical imaging. Non-invasive functional testing can be used to diagnose obstructive CAD by detecting ischemia through ECG changes, wall motion abnormalities by stress CMR or stress echocardiography and perfusion changes by SPECT, PET, myocardial contrast echocardiography or contrast CMR. These tests have a high accuracy for detecting flow-limiting coronary stenosis. (15) On the other hand, low-grade coronary stenoses not linked to ischemia, remain undetected with these techniques. Non-invasive anatomical assessment can be done with coronary CT angiography (CCTA),

in which an intravenous contrast agent is used to visualize the lumen of the coronary arteries. This technique also has a high accuracy in detecting obstructive CAD. (15) Stenoses with an estimated occlusion of 50-90% by visual inspection should be evaluated with non-invasive or invasive functional testing, as these stenoses do not always induce ischemia. (15, 16)

Invasive testing

ICA is the gold standard for the diagnosis of CAD, as well as for intraprocedural guidance of percutaneous coronary interventions (PCI). (17) However, diagnostic ICA is only indicated in patients suspected of CAD, when non-invasive testing is inconclusive or when non-invasive testing suggests a high event risk. (17) Invasive functional assessment can be done during ICA. This is especially important in patients with intermediary stenoses (50-90% occluded) and multivessel disease, as there can be a mismatch between anatomical and functional severity of stenoses. (18) This can be done with the instantaneous wave-free ratio (iFR) and the fractional flow reserve (FFR). (19) In iFR a pressure wire is used to measure aortic blood pressure and the blood pressure distal to the stenosis under investigation. When the calculated ratio lies below 0.89, a stenosis is typically classified as significant. (20) FFR is determined by the ratio of the mean distal coronary artery pressure to the means aortic pressure during maximum hyperemia. A stenosis is classified as significant when this ratio drops below 0.80. (21)

1.1.4 Treatment of CAD

The treatment of CAD consists of lifestyle changes, optimal medical therapy (OMT) and interventions when necessary. The goal of the pharmacological therapy is to reduce symptoms and prevent cardiovascular events. Interventions are used for revascularization and are considered when medical therapy is insufficient. (22) The objectives of revascularization are symptom relief and improvement of the prognosis. Revascularization can be done via a PCI or coronary artery bypass grafting (CABG). In PCI stents are placed in stenosed spots in the coronary arteries. This is done via heart catheterization. A stent is brought into position under fluoroscopic guidance and is then expanded with a balloon to counter the narrowing. A schematic representation of this process is shown in figure 1.3.

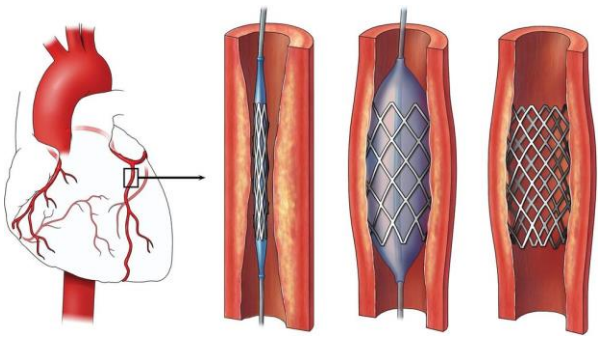


Fig. 1.3 PCI: a wire with a balloon and a stent is placed at the desired location in the stenosis. By inflating the balloon the stent is expanded and anchored in the artery. (23)

1.2 Technical background

1.2.1 Artificial intelligence

The definition of artificial intelligence (AI) is very broad and ever changing. Any intelligent behavior displayed by machines can be seen as AI. More recent definitions are built around imitating intelligent human behavior. (24) AI in medicine can generally be divided in virtual and physical applications. Virtual applications range from prioritizing radiology worklists to intraprocedural guidance of interventions. Physical applications include robot assisted surgery, advanced protheses and elderly care. (25)

1.2.2 Machine learning

In evidence-based medicine, guidelines are developed based on correlations and associations found in clinical data. This is why a specific subset of AI, machine learning (ML), is increasingly popular in medicine. (26) In ML, mathematical algorithms perform intelligent predictions based on datasets and improve through experience. (27) The development of a ML algorithm starts with a training phase, in which a large set of data is used to learn the algorithm what features can be used to predict to which class each input belongs. This dataset consists of examples coupled with a class, also called a label. By looking at all the different examples, the algorithm can learn what parameters are predictive for each label. This means that after the training phase, the algorithm can predict the label for new unseen data. This is called supervised learning, as each example is coupled to a label. In unsupervised learning, all data is given to an algorithm without a label. The algorithm itself can then start to try grouping similar examples together. This is called unsupervised learning and in this way new relations between parameters can be discovered. (28)

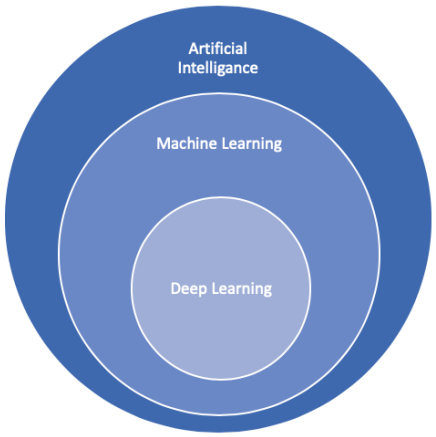


Fig. 1.4 Relation between artificial intelligence, machine learning and deep learning. (29)

1.2.3 Deep learning

Deep learning (DL), a specific subset of ML, has especially gained interest over the last few years. The main reason for this is that DL can find correlations that are too complex for other algorithms. In other ML algorithms feature engineering is needed for the construction of an algorithm. For feature engineering domain expertise and human choices are needed to create algorithms that transform raw data, such as an X-ray image, into variables in which a ML algorithm can detect patterns. In DL, an algorithm can be given raw data and it creates its own representation that can be used for pattern recognition. The difference between DL and other ML techniques is shown in figure 1.5. This is done in multiple layers, each layer with its own representation of the data. Typically, the data moves through several layers sequentially, with each layer transforming the data into more abstract representations. In this way, a DL algorithm can learn highly complex functions. (30) Due to this, DL is well suited for image analysis and many studies have already shown promising results. (30) Convolutional neural

networks (CNNs), a type of DL algorithm, are the state-of-the-art in image analysis and outperform humans in several image analysis tasks. (31, 32)

Over the last years, ML has shown that it can potentially revolutionize healthcare in several ways. By improving diagnostics, therapy decision making, logistic optimization and changing the healthcare model from a reactive to a more proactive approach by predicting outcomes and adverse events. (33)

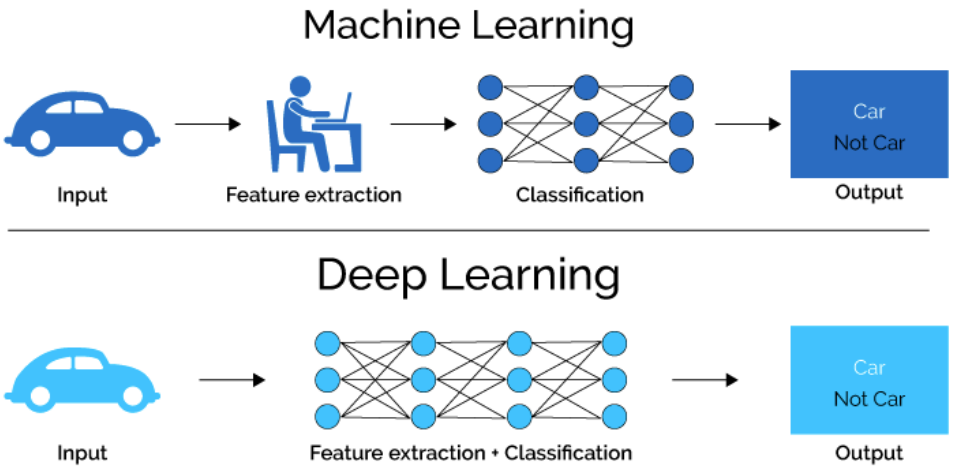


Fig. 1.5 Difference between machine learning and deep learning. In DL no feature extraction is needed. (29)

1.3 Problem definition

As mentioned before, ICA is the gold standard for diagnosing CAD. In ICA the lumen of the coronary arteries is assessed by injecting them with radiopaque contrast. (17) To fully assess all coronary arteries, multiple projections should be used. The coronary arteries are a 3D structure and the images acquired are 2D projections. Because of this, not all arteries are clearly visible from all angles, as they might appear to overlap in some projections. Furthermore, when using only one projection, some stenoses can be missed. If an eccentric stenosis is present, the arteries can appear normal or not significantly narrowed form one view, while it appears significant in another projection. (34) This can be seen in figure 1.6. When looking from B, the stenosis looks insignificant, as most of the plaque in located inline with the projection. When looking from A, the stenosis appears significant as the plaque is positioned orthogonal on the projection.

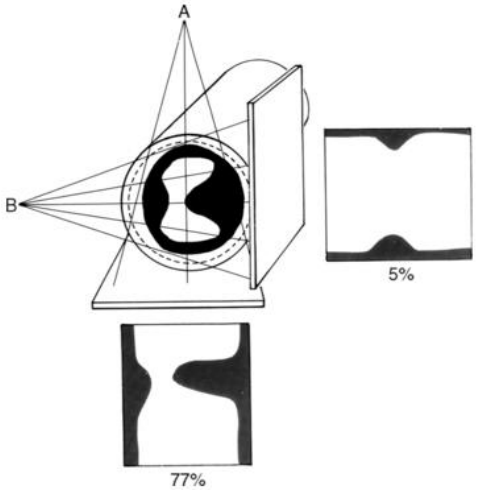


Fig. 1.6 Schematic representation of how eccentric stenoses can appear insignificant in one projection and significant in another. (35)

To prevent errors in stenosis assessment and assess all coronary arteries properly, multiple projections are used. The projections are named based on the position of the image intensifier. In the anteroposterior (AP) projection the image intensifier is positioned right above the patient. The X-ray tube can be angulated horizontally. In the right anterior oblique (RAO) projection the image intensifier is moved towards the right shoulder of the patient. In the left anterior oblique (LAO) the image intensifier is moved towards the left shoulder of the patient. The RAO and LAO positions are shown in figure 1.7. Vertical angulation is also possible, resulting in a cranial and a caudal position, also shown in figure 1.7. Vertical and horizontal angulation can be combined to create different projection of the coronary arteries. An overview is shown in table 1.1.

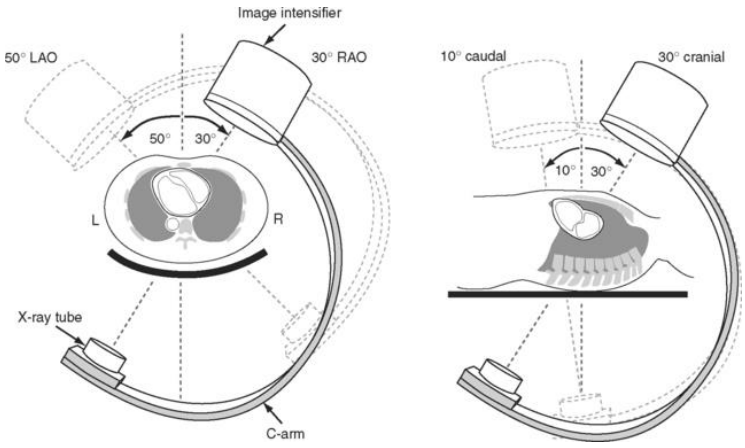


Fig. 1.7 Possible positions of the X-ray tube in coronary angiography. In the left image horizontal angulation is shown. In the right images vertical angulation is shown. (36)

Because coronary artery anatomy can differ over patients, the optimal angle for visualization can also differ. However, generally cranial views are used to visualize the LAD, caudal views for the RCx and LAO, RAO and LSO for the RCA.

Table 1.1 Different types of views in ICA

		Vertical angulation		
		AP	**Cranial**	**Caudal**
Horizontal angulation	**AP**	AP	Cranial	Caudal
	Right	RAO	RSO	RIO
	Left	LAO	LSO	LIO

AP: Anteroposterior view, RAO: Right anterior oblique, LAO: Left anterior oblique, RSO: Right superior oblique, LSO: Left superior oblique, RIO: Right inferior oblique, LIO: Left inferior oblique.

At this moment, stenosis severity is determined via visual inspection by a cardiologist. This method has several important drawbacks: a significant inter- and intra-rater variability and a high positive prediction bias. This positive prediction bias leads to unnecessary PCI's, exposing patients to a risk of complications while not necessary. (37) This is especially evident in cases with intermediary stenoses, where lesion may appear significant in one view and not in the other. (38) To overcome these issues, intracoronary functional measurement and imaging were developed. Functional assessment can be done by determining iFR or FFR values, based on invasively measuring the difference in blood pressure in the aorta and in the coronary artery distally of the stenosis under investigation. (19, 21) In intracoronary imaging (ICI), a catheter for the acquisition of intravascular ultrasound images or optical coherence tomography images is inserted and moved through the coronary arteries. These techniques improve the accuracy of stenosis assessment, however, procedural time is prolonged and complications risks and costs are increased. (39, 40) Even with the combination of ICA, functional measurements and ICI, a relevant number of syndromes exists, that cannot be diagnosed sufficiently. (37, 41) This can generally present in two ways, either stable, called ischemia with non-obstructed coronary arteries (INOCA) or as a myocardial infarction called myocardial infraction without obstructed coronary arteries (MINOCA). (42)

1.3.1 Ischemia with non-obstructed coronary arteries

Patients suffering from INOCA often present with typical ischemia symptoms and a limited coronary flow reserve (CFR). For these reasons patients are referred for ICA. However, no obstructive CAD (≥50% diameter stenosis) is then found. (43) Almost two thirds of woman and one third of men with stable AP symptoms undergoing ICA do not have obstructive CAD. Furthermore, this presentation is associated with an increased risk of major cardiovascular events (MACE), persistence of symptoms and a higher re-hospitalization and re-catheterization rates compared to asymptomatic individuals. (42, 44) The incidence of adverse events and impaired quality of life are comparable to patients with obstructive CAD. (45) Due to the absence of obstructions on ICA images, INOCA is often not diagnosed and hence no specific therapy is given. Ischemia can objectively be proven with non-invasive techniques such as (contrast) echocardiography, PET, CMR and SPECT. However, it should be noted that SPECT and stress echocardiography cannot diagnose INOCA in certain cases where the whole left ventricle is affected. (46, 47) The etiology of INOCA is heterogeneous and includes coronary vasospasm and microvascular dysfunction. In most cases, microvascular dysfunction is caused by either structural microcirculatory remodeling, functional arteriolar dysfunction or a combination of both. (48) Structural remodeling is caused by inward remodeling of coronary arterioles with an increased wall to lumen ratio, loss of capillary density or both. This can be caused by cardiovascular risk factors, atherosclerosis, left ventricular hypertrophy or cardiomyopathy. (49) This results in a decreased vasodilatory range, limiting oxygen supply to the myocardium. Furthermore, there is an increased sensitivity to vasoconstricting stimuli. (50) This can be found as reduced CFR under adenosine testing and an

increased minimal microcirculatory resistance. In the arteriolar dysfunction, the medium to large sized arterioles have limited vasodilation capabilities, caused by endothelium dysregulation. (51) There might even be paradoxical vasoconstriction with increased myocardial oxygen consumption. This can be found as a limited increase of blood flow under acetylcholine challenge (<1.5 times resting flow), a reduction in blood flow due to arteriolar spasms without epicardial vasospasm and diffuse narrowing of distal epicardial vessels without focal coronary spasm. (45) Epicardial vasospasms generally originate from a hyperreactive epicardial coronary segment, undergoing maximal vasoconstriction when exposed to a vasoconstrictor stimulus. (52) Stimuli can be smoking, emotional stress, drugs, peaks in blood pressure, allergic reaction, and the implantation of stents. (53)

In the 2020 expert consensus document by the European association of percutaneous cardiovascular interventions (EAPCI) and European society of cardiology (ESC), five INOCA endotypes are recognized. These are microvascular angina (MVA), vasospastic angina (VSA), combined MVA and VSA, non-cardiac chest pain and non-flow limiting CAD, caused by diffuse coronary artery atherosclerosis. (45) These types can be distinguished by functional testing during ICA. First, standard ICA images are acquired, to assess the presence of obstructive CAD. Then hyperemic testing can be done based on adenosine to determine FFR. CFR can be calculated based on thermodilution or doppler flow velocity. (54, 55) The index of microvascular resistance (IMR) can be determined by combining the flow and pressure measurements. (56) If the FFR lies below 0.8, there is a significant stenosis present. When the CFR lies below 2.0 and the IMR ≥25, there is coronary microvascular dysfunction. The next step is to determine vasoreactivity via acetylcholine (ACH) testing. The test is considered positive if intracoronary ACH infusion results in anginal symptoms, ischemic ECG changes and ≥ 90% reduction of epicardial vessel diameter. (57) The combination of the above-mentioned tests makes it possible to divide patients in the four endotypes. An overview of this process in shown in figure 1.8.

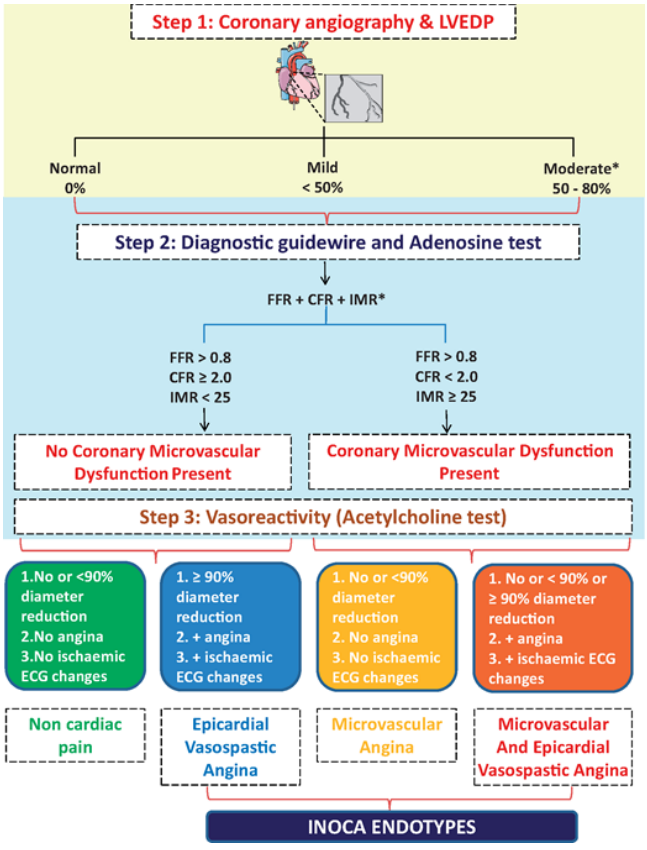


Fig. 1.8 Different INOCA endotypes and how to distinguish them. (45)

Treatment of patients with INOCA is challenging, as it is a heterogenous group and randomized trials are lacking. (45) Generally, it consists of advice on lifestyle factors, risk factor management and antianginal medication. It is known that INOCA is often not correctly identified and that this leads to poor outcomes for patients. Furthermore, the diagnostic process described above is extensive. This shows the need for more diagnostic tools to improve management of this patient group. (45)

1.3.2 Myocardial infarction with non-obstructed coronary arteries

There is a significant group of patients presenting with all signs of myocardial infraction, but without obstructive CAD. Up to 14% of patients with an acute myocardial infarction (AMI) does not have obstructive CAD. (58) MINOCA is defined as the presence of an AMI (based on the 4th universal definition of AMI, shown in table 1.2), non-obstructive coronary arteries on ICA (defined as no coronary stenosis ≥50% in any potential infract-related artery) and no clinically specific cause for the acute presentation. (59) Approximately, one third of patients presenting with MINOCA has ST-segment elevation on their ECG. (58) A meta-analysis by Pasupathy et al. (2015) showed that MINOCA patients have a better 12-month all-cause mortality compared to patients with obstructive myocardial infraction (4.7% vs.6.7%). However, 12-month mortality is significantly higher than in patients with stable AP (4.7% vs. 0.3%) (58) So, MINOCA seems to have a clearly worse prognosis than stable AP. According to the ESC, MINOCA should be considered as a working diagnosis. (60) This means that underlying causes have to be investigated. The differential diagnosis for MINOCA consists of plaque disruption, coronary artery spasm, coronary thromboembolism, coronary dissection, Takotsubo cardiomyopathy, myocarditis and type 2 AMI. (60)

Table 1.2 Definition of acute myocardial infarction. (61)

The fourth universal definition of acute myocardial infarction (AMI) defines AMI as the presence of:
1. Acute myocardial injury with clinical evidence of acute myocardial ischemia,
2. Detection of a rise and/or fall of cardiac troponin with at least one value above the 99th percentile upper reference limit
3. With at least one of the following: - Symptoms of myocardial ischemia - New ischemic ECG changes - Development of pathological Q waves - Imaging evidence of new loss of viable myocardium or regional wall motion abnormality in a pattern consistent with an ischemic etiology - the identification of a coronary thrombus by angiography or autopsy

1.3.3 Deep learning as a possible solution

As described above, ICA has several drawbacks and automated software that uses all information within the ICA images and relates it to the context of complaints and outcomes, could be a valuable tool for improving the diagnostic and therapeutic process. Since ML can find relations between patient groups based on images, this could be a possible solution. (28) ML is essentially different from image analysis by physicians, which is associated with several drawbacks as described earlier. (37, 38) Furthermore, it has been shown in several other imaging modalities that ML-based algorithm can be used to predict disease prognosis. (62) In a study by Bentacur et al. (2018), a deep learning model was developed that could predict major adverse cardiac events based on SPECT myocardial perfusion imaging (MPI) better than experts and standard quantifications. (63) If these kinds of algorithms could be used during ICA acquisition, it could improve diagnostics and guide treatment decision making.

However, despite promising results in both radiology and cardiology, implementation of ML is still limited. (62, 64) There are several reasons for this limited implementation. Current ML research is often not really focused on clinical practice and patient outcomes. Often metrics do not represent outcomes for patients and algorithms are difficult to compare. This makes it hard for clinicians to identify where ML can improve healthcare and what the actual value of these algorithms is. (65) Lack of knowledge on ML is also a factor that limits implementation in clinical practice. (66) There are also several factors intrinsic to ML that play a role in this. These include a lack of generalizability, bias and fitting on confounders in the used dataset. (66) These problems are all related to the amount and quality of the data used for training. Availability of large high quality datasets can resolve or reduce

these problems and so increase the positive impact of ML on healthcare. (64) This shows that there is a need for a systematic approach for the development and implementation of ML-based algorithms to improve healthcare and use ML to its fullest potential. As availability of sufficient and high-quality data is a cornerstone for the development of well performing ML models, focus should be on creating datasets that meet these requirements.

1.4 Thesis overview

1.4.1 Thesis motivation and outline

Over the last 15 years imaging data acquired during ICA in the cardiology department of the Leiden University Medical Center (LUMC) was digitally stored. These images can be related to the electronic health records (EHRs) via the patient identification number and date of cardiac catheterization. This means that there is a very large amount of data available. Utilizing this large amount of data in a correct way can lead to innovations and improvement of patient care and management.

To develop ML-based algorithms that really can improve healthcare, a thorough curation of data for training is necessary. Furthermore, a focus on clinical practice is essential. This requires knowledge of both the medical and the technical background. In addition to that, bringing expert on each topic together and understanding both sides of the spectrum is important. These things make a technical physician very well suited for this process.

The aim of this project is to create an overview of what is needed for the creation of datasets suitable for the training of ML models in the clinical practice of interventional cardiology and how this can be done. This was done in several steps. First a literature study was carried out to assess the current state of AI-based ICA image analysis in interventional cardiology. This review can be found in Chapter 2. Next the expectations and perceived barriers of interventional cardiologists towards AI and possible applications of AI in interventional cardiology were assessed, by carrying out structured interviews. The results can be found in Chapter 3. Following to that, the feasibility of querying EHRs based on health insurance codes for the creation of a MINOCA dataset was assessed. This was done by carrying out several queries and determining the actual status of the patients found by the query, by analyzing the information in the EHRs manually. This can be found in Chapter 4. After this, a proof-of-concept study was carried out in which a deep learning model was developed to predict hemodynamic significance of coronary artery stenoses on ICA images. This was done to create an overview of the steps needed for creating a dataset and a model and not to create a generalizable model for clinical practice. Furthermore, this gives an insight in the logistics and infrastructure needed for this type of research. This can be found in Chapter 5. Finally, a roadmap for the creation of datasets for ML models in the clinical practice of interventional cardiology was created. This was based on all insights gained during the other steps described above combined with scientific literature. This can be found in Chapter 6.

1.4.2 Central research questions

Central clinical question:
Is it possible to use ML to solve diagnostic challenges such as encountered in MINOCA?

Central technical question:
What is needed to apply ML on ICA images and other sources of coupled medical data?

1.4.3 Research goals

1. Assessing the expectations and perceived barriers by interventional cardiologists on ML-based algorithms in clinical practice.
2. Assess querying methods with an EHR system for the creation of a ML dataset.
3. Conducting a proof-of-concept study on predicting lesion significance on ICA images.
4. Creating a roadmap for the curation of data for the development of ML models in interventional cardiology.

2. The current state of artificial intelligence based invasive coronary angiography image analysis: a literature review

2.1 Introduction

Coronary artery disease (CAD) is one of the leading causes of death worldwide and the incidence is still increasing. (5) Invasive coronary angiography (ICA) is the gold standard for the diagnosis of CAD as well as for intraprocedural guidance of percutaneous coronary interventions (PCI). (17) In ICA the coronary arteries are imaged by injecting them with radiopaque contrast, while a continuous X-ray recording is made to assess the lumen of the coronary arteries. At this moment, stenosis severity is typically determined by visual inspection by a cardiologist. This method has several important drawbacks: a significant inter- and intra-rater variability, a high positive prediction bias and the inability to diagnose syndromes with myocardial ischemia without significant stenoses on ICA images. (37, 41) Furthermore, a relevant number of ischemic syndromes without clear stenoses on ICA exists and can be up to 50% of all patients undergoing ICA. (41) Examples of these syndromes are: coronary microvascular disease (CMD), myocardial infarction with non-obstructed coronary arteries (MINOCA) and generalized atherosclerosis without focal stenoses. (67)

A current solution to the subjective nature of visual inspection in ICA is the use of quantitative coronary angiography (QCA) software. In QCA, ICA images are semi-automatically analysed, resulting in a lower positive prediction bias and less variability. (68) However, subjective choices are still needed in this procedure. QCA is still not widely used in ICA as it is time-consuming nature does not fit in the cardiac catheterization workflow. (37, 68) In addition to that, QCA is not able to overcome the diagnosis issue of ischemic heart disease with non-obstructed coronary arteries. A currently used approach for solving this problem is the use of intracoronary imaging, such as intravascular ultrasound (IVUS) and optical coherence tomography (OCT), and functional assessment such as fractional flow reserve (FFR) or instant wave-free ratio (iFR). However, these types of assessment prolong procedural time, increase complication risks for patients and make use of costly materials. (39, 40)

Automated software that uses all information within the ICA imaging dataset and relates that to the context of complaints and outcomes, could be a valuable tool within the current workflow. Since artificial intelligence (AI) can find relations between patient groups based on images, this could be a possible solution. As it is possible that the images hold more information than meets the eye. A study by Kwak et al. (2020) in patients with aortic stenosis demonstrated that AI can be used to identify outcome types based on echocardiography results, clinical characteristics and laboratory data. (69)

Machine learning, a subfield of AI in which computer algorithms are created that allow computer programs to automatically improve through experience, is a technique that has potential for the automated analysis of ICA images. In machine learning mathematical algorithms perform intelligent predictions based on datasets. (27) Deep learning algorithms, a subset of machine learning, is especially well suited for image analysis. Deep learning has not only surpassed other machine learning algorithms in image analysis, but also outperforms humans in several tasks. (31, 32) The main difference between deep learning and other machine learning algorithms, is that for deep learning no pre-programmed instructions on which data features to use are needed. This means that previously unknown relations can be discovered and can be used to solve tasks like localizing and classifying objects in images. (70)

In this literature review the current state of machine learning based algorithms in ICA image analysis was investigated. It is focused on image analysis by machine learning algorithms, rather than report

and text analysis, as the incorporation of machine learning algorithms in the cardiac catheterization workflow is especially of interest. Within this review, the exact purpose of algorithms, used AI techniques, characteristics of datasets, accessibility of algorithms and performance are discussed.

2.2 Methods

Eligible studies were identified by searching the following electronic databases: Medline, EMBASE, Web of Science, IEEE, and Google scholar. Additional literature was identified by checking references of included literature and by a search on ArXiv. The final search was performed on September 21st, 2020. The search strategies for each database can be found in Appendix 1.

After duplicates were removed, eligibility was assessed based on title, abstract and full text. First a global screening was performed based on title and abstract. To be eligible articles had to scientific papers written in English and about direct ICA image analysis by AI algorithms. For the list of inclusion criteria see table 2.1. This means that articles about machine learning based analysis of electronic health records and catheterization reports were excluded. After global screening, the not excluded articles were reviewed more in depth on full text.

For every included article the following items were reviewed: the purpose of the developed algorithm, type of algorithms developed, data characteristics and annotations used, algorithm performance and the accessibility of the used data and the source code.

Table 2.1. Inclusion criteria for reviewed literature

- Scientific paper
- Written in English
- Invasive coronary angiography as main imaging modality
- Research on artificial intelligence algorithms
- Algorithm for image analysis

2.3 Results

Study selection

By applying the different search strategies to all the databases, 1683 articles were collected. After removing duplicates 1429 articles remained. Next, all titles and, if appropriate, abstracts were screened. This resulted in the exclusion of 1311 articles. Out of all articles, 52 were excluded because they were books, 28 articles were not written in English, 679 articles were not about ICA, 534 articles did not perform research on AI algorithms and 18 articles did not perform image analysis. The remaining 118 articles were assessed for eligibility by reviewing the full text, if available. This led to the exclusion of another 80 articles: 45 articles were not about ICA, 19 articles did not research AI algorithms and 16 articles did not perform image analysis. The whole process of inclusion is shown in figure 2.1.

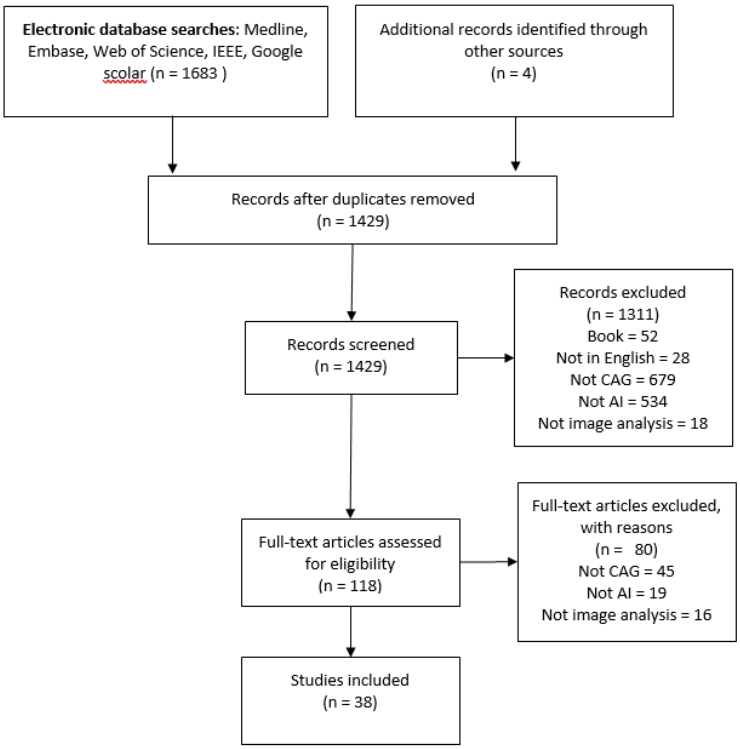


Fig. 2.1 PRISMA flow diagram for inclusion of studies.

Study characteristics

For all included studies, the characteristics are presented in table 2.2-4. The studies were divided into three groups based on the general goal of the developed algorithm. Table 2.2 shows all characteristics for the 18 studies focused on *segmentation* tasks. Table 2.3 for the 15 studies on *classification* tasks and table 2.4 for the 5 studies focused on *other* tasks.

Segmentation studies

In the *segmentation* literature, algorithms for segmentation tasks were researched. Segmentation is here defined as the task of dividing all pixels in a picture into groups, in this case into two groups: *background* and *coronary arteries*. The general approach of most studies was quite similar. First, a pre-processing step, which varies between studies, was applied. Pre-processing was done to prepare the images for analyses by the developed algorithm. This is followed by the segmentation step. The choices made by the algorithm in the segmentation step are based on the data used for training, which are ICA images and their segmentation ground truth. The ground truth can be constructed in several ways. The

most common approaches were manual segmentation by experts or semi-automatic segmentation with manual correction by experts. In figure 2.2 a standard ICA image, a segmentation ground truth and an algorithm segmentation output are shown.

An overview of all study characteristics is shown in table 2.2. The year of publication for the included segmentation studies ranges for 1997 to 2020, with one study published in 1997, one in 2012 and the others between 2016 – 2020. Various machine learning algorithms were researched in this group. However, 18 out of 19 studies used a deep learning approach. The most frequently used deep learning algorithm is a convolutional neural network (CNN), which was used in 10 studies. Four studies used an artificial neural network (ANN), one study used a generative adversarial network (GAN), one study used a fully convolutional network (FCN) and one study used a Hopfield network. Plourde et al. (2012) was the only study not using a deep learning network, instead a support vector machine (SVM) was used. In all studies ICA images were used as a training dataset. There is a range in the amount of data used: from 2 ICA images to 727 full ICA sequences, where a sequence typically consists of 15-25 images. Furthermore, several studies used additional non-ICA data. Yu et al. (2019) used the DRIVE dataset for transfer learning, this dataset consists of 40 fundus images with annotation in the form of vessel segmentation. Ma et al. (2019) used the flying chair dataset. This is a synthetic dataset with optical flow ground truth, consisting of 22872 image pairs and corresponding flow fields. Plourde et al. (2012) used the CAVAREV dataset. This data set contains angiographic phantom images in the form of two simulated dynamic projections on a phantom with contrasted coronary arteries, derived from patient data. For data annotation most studies used manual segmentation. In 14 studies the ICA images were manually segmented, in three studies the ICA images were semi-automatically annotated and manually corrected. Yu et al. (2019) did not annotate the ICA data, however, the fundus images where manually segmented. The availability of the used datasets and source code were not mentioned in 14 studies. Cervantes-Sanchez et al. (2019) mentioned that the annotated database is available online. The source code and dataset for Hao et al. (2020) are both available online. The before mentioned DRIVE, flying chair and CAVAREV dataset are also available online.

Fig. 2.2 From left to right: standard ICA image, manually segmented ground truth and algorithm output. (71)

Classification studies

In the *classification* literature, algorithms were developed that can divide images, or parts of images into different categories, based on the local features of that image. This can be applied in several different ways to ICA images. Possible applications are stenosis detection, stenosis classification (lesion type, morphology, significance), dividing coronary arteries into anatomical segments and predicting patient related parameters. figure 2.3 shows an example of how such an algorithm can be implemented.

An overview of all study characteristics is shown in table 2.3. For all included classification studies, the year of publication ranges from 2014 to 2020, with one article published in 2014 and the remainder in 2018-2020. Various types of algorithms were developed. In 13 out of 15 studies, a deep learning algorithm was developed. Besides this, Hae et al. (2018) developed multiple types of algorithms and Cho et al. (2018) constructed a gradient boosting algorithm. The most frequently used deep learning algorithm was a CNN. This approach was used in 12 studies. In one of these studies a GAN was combined with a CNN. Kumar et al. (2014) developed an ANN. Hae et al. (2018) tested a K-nearest neighbour, binary class L2 penalizes logistic regression, SVM, random forest, extra tree, AdaBoost, light GBM, CatBoost, Gaussian naïve, Bayes and an ANN algorithm. In all studies ICA images were used for training. The amount used ranges from 50 to 675000 ICA images. Besides this, several studies used additional data. Cho et al. (2019) used Fractional flow reserve (FFR) assessment, age, Body surface area (BSA) and the notion of the involved segment, besides ICA images. Hae et al. (2018) additionally used FFR assessment, coronary CT angiography (CCTA) and intravascular ultrasound (IVUS) data. Data annotation was generally more extensive for classification algorithms as compared to segmentation algorithms, as additional information might be needed to divide images into distinct groups. Segmentation is often used as annotation and this is done manually, semi-automatically with manual correction or fully automated. Besides this, several studies also annotated each vessel with the corresponding anatomical term. Other annotations used on vessel level are centreline, lumen diameter, diseased vs. non-diseased, minimal lumen diameter (MLD), bifurcation label and reference vessel diameter (RVD)/reference lumen diameter (RLD). Stenosis annotation was done either manual or semi-automatically with QCA software. Annotations used for stenoses labelling were location, morphology, type of lesion, lesion length and significance. The availability of the datasets, source code or training details were not mentioned in 11 studies. Cho et al. (2019) explicitly mentioned that this will not be made publicly available. For two studies the training details are available, but the other aspects were not mentioned. Lee et al. (2019) collected all used data via a web crawler and so used only publicly available data. A web crawler is a computer program that downloads and indexes specific content all over the internet to learn what information is available and where to find it. Hae et al. (2018) mentions that all used data, source code and calculated features are publicly available.

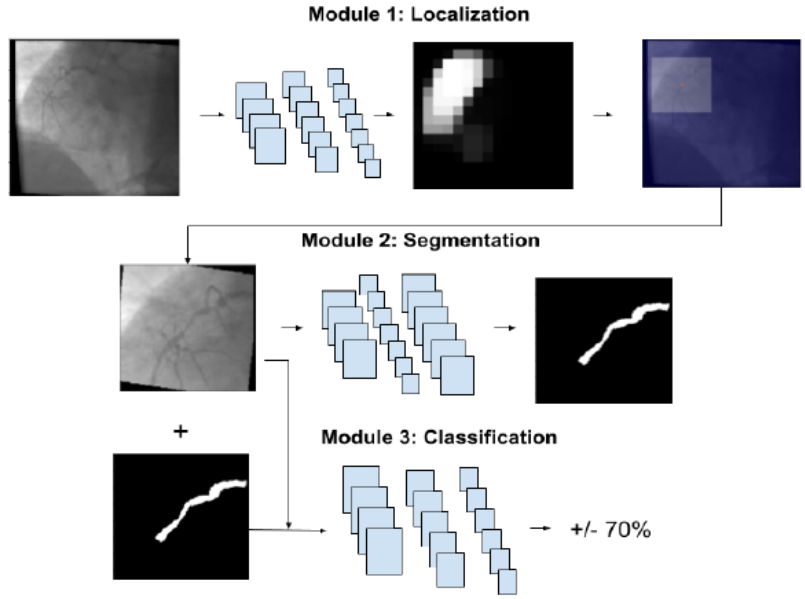


Fig. 2.3 An implementation of a classification algorithm. First a classifier localizes a possible stenosis, then it is segmented and finally it is classified as significant or not. (72)

Other studies

Five studies that were found were neither on segmentation nor on classification. In these studies, several different tasks were performed, some of which might be of additional value for segmentation and classification tasks. The tasks were catheter detection, cardiac phase detection, image matching, contrast inflow detection and vascular tracking.

A catheter detection algorithm is able to discriminate between the catheter and vessels in an automated way, as this is not always clear by appearance. (73) Catheter interference is a frequent cause of artefacts in automated ICA image analysis. (74) The cardiac phase detection algorithm was trained by coupling ECG data with the ICA images. In this way the algorithm can determine in which phase of the cardiac cycle the heart is, by analysing the ICA images. This can be used to select images on which the vessels are best visible and calculation on flow can be made. (75) The Image matching algorithm can match images from different angles and different moments in time with each other. This is done by calculating multilevel correlation maps and matching these. This technique can be used to follow a periodic motion and to generate 3D reconstructions of the heart. (76) A contrast inflow detection algorithm is used to detect the start of contrast flow in the coronary arteries. By analysing the average pixel intensity, the algorithm can detect the start. This can be useful for frame selection for further analysis of ICA sequences. (77) A vascular tracking algorithm can be used to track the position of coronary arteries over time. Every branch is given a class label and each branch is followed over time. After this the separate branches are merged to one structure again. This technique can be used to analyse continuity over time, create a 3D reconstruction over time and assess dynamic information. (78)

An overview of all study characteristics is shown in table 2.4. All studies were published between 2012 and 2020. Four out of five used deep learning algorithms. Three out of these four used a CNN approach. Ma et al. (2017) used both a CNN algorithm and a long-short term memory recurrent neural network (LSTM-RNN) and compared these two. Hernandez-Vela et al. (2012) combined an Adaboost algorithm with multiscale stacked sequential learning for catheter detection. In all studies ICA images were used, the amount of data ranged from 50 images pairs to 45700 full ICA sequences. Ciusdel et al. (2020) additionally used ECG recordings coupled to the ICA images. Fan et al. (2019) used the KITTI optical flow dataset, consisting of 3000 image pairs. Data annotations was very divers in this group, as they focus on different applications. Hernandez-Vela et al. (2012) focused on catheter detection and data annotation consisted of manual catheter annotation. Ciusdel et al. (2020) developed an algorithm for cardiac phase detection and annotation consisted of diastolic – systolic phase switch annotation on the ECGs, which was done by an automated signal processing algorithm. A subset of ECGs was manually annotated by experts. Fan et al. (2019) developed an algorithm for image matching. The annotation of the used KITTI database consisted of matching points between the image pairs. For the ICA data each image pair was manually annotated with 100 correspondence points. Ma et al. (2017) developed an algorithm for contrast inflow detection and each ICA sequences was manually marked for the start of contrast inflow. Fang et al. (2018) developed an algorithm for vascular tracking. The data annotation consisted of centrelines for all vessels. This was manually done by three experts and the average value of all three was used. Hernandez-Vela et al. (2012) made the used dataset available online. Ciusdel et al. (2020) explicitly mentioned that the algorithm will not be made commercially available. The KITTI dataset used by Fan et al. (2019) is available online. The availability of the used data and source code of the other studies was not mentioned.

Algorithm performance

An overview of the algorithm performance per article is shown in table 2.5-7. Table 2.5 shows summarizes the segmentation literature, table 2.6 the classification literature and table 2.7 the other literature. The tables show on what kind of dataset the metrics were calculated, the size of the used datasets, on which algorithm the metrics were calculated and the calculated values for the used metrics. There are generally three possible types of datasets to calculate algorithm performance on. These are a training set, a validation set and a testing set. The training set contains the data used for learning, on which the algorithm parameters are fitted. The validation contains data used to tune the parameters of the algorithms, i.e., number of layers in a neural network. The test set contains data that is only used to assess the performance of a fully specified algorithm. (79) There are two possibilities for the test set. It either contains data from the same source as the training and the validation set or an external dataset can be used. An external dataset comes from a different source than the training and validation set.

It is clearly visible in the tables that many different metrics was used to assess algorithm performance. However, most of these metrics are related, as they are based on the number of false positives, true positives, false negatives and true negatives. Furthermore, the part of the developed algorithm for which the metrics are calculated differs between studies. Some authors calculate it on the whole end-to-end algorithm, however others only on parts of it.

Table 2.5 shows the performance metrics of the *segmentation* literature. Only Yu et al. (2019) calculated performance on the validation set. Yang et al. (2019, (80)) and Karapataki et al. (1997) used an external set to evaluate algorithm performance. In all other cases a test set was used. The size of the dataset used for performance evaluation ranges from 1 to 1788 ICA images. In most studies, the whole proposed algorithm was evaluated end-to-end, so all steps are considered, such as frame selection modules and image pre-processing. However, Yang et al. (2018, (80)), Yang et al. (2019, (81)), Hao et al. (2020) and Jun et al. (2020) evaluated multiple algorithms for the same application. For these cases, the performance of the best performing algorithm is shown. In the study by Yang et al. (2018, (80)), the DenseNet121 based method was used. In this study, the algorithm was applied in a per vessel and a whole vascular tree strategy. The results for the whole vascular tree are shown, as this is most similar to clinical practice. For Hao et al. (2020) the 3D + CAB algorithm was the best performing. For Jun et al. (2020) the optimized T-net based method was the best. The performance on the whole vascular tree is shown here. Jo et al. (2018) presented the performance based on the average values of all evaluated selective feature mapping algorithms. For every metric that is used more than once, the highest value is presented in bold. Fan et al. (74) found the highest accuracy and specificity.. Nasr-Esfahani et al. (2016) had the highest precision, Yang et al. (2019, (80)) the highest dice similarity coefficient (DSC) and the highest sensitivity. Karapataki et al. (1997) does not mention any performance metrics, however computational time is discussed.

Table 2.6 summarizes the performance metrics of the *classification* literature. Lessage et al. (2020) and Liu et al. (2019, (82)) calculated performance on a validation set, note that Liu et al. (2019, (83)) evaluated two algorithms with two distinct goals and because of this it is shown two times in the table, similar to the studies of Du et al. (2020), Lee et al. (2019) and Liu et al. (2019, (83)). Cho et al. (2019) and Hae et al. (2018) used an external dataset for testing. Cong et al. (2019) and Zhang et al. (2019) used 4-fold and 10-fold cross-validation, respectively. All other authors calculated performance metrics on a test set. The size of the dataset used for evaluation ranges from 50 to 135000 images. In several studies, multiple algorithms were independently evaluated. Cong et al. (2019) evaluated an algorithm classifying the stenoses into two groups (<25% stenosis or >25%) and another algorithm with

three possible classes (<25% stenosis, 25-99% stenosis and total occlusion). The results of two-group-analysis are shown. Du et al. (2020) developed an algorithm classifying parts of the vessels into different segments and an algorithm classifying the morphology of stenoses. The values shown for morphology classification are an average of the calculated metrics per morphology class. Lee et al. (2019) also evaluated two algorithms, an algorithm labeling all parts of the coronary arteries with their anatomical name and a stenosis detection algorithm. Liu et al. (2019, (83)) also evaluated a stenosis detection and a morphology classification algorithm. Hae et al. (2018) evaluated multiple algorithms for the same goal, the one with the best performance was the Catboost based algorithm, for which the performance is shown. In the article by Liu et al. (2019, (82)) a bifurcation detection algorithm and a lesion detection algorithm are evaluated. Yang et al (2019, (84)) also develop two algorithms, however, only the performance of vessel classification was calculated and not for the stenosis classification. Therefore, only the first is shown. As the goal of the different algorithm varies, it is not possible to determine which performs the best.

Table 2.7 shows the performance metrics of the algorithms with a purpose *other* than segmentation of classification. Ciusdel et al. (2020) used an external dataset to evaluate the algorithm performance. Fan et al. (2019) used a test set. In the other articles a validation set was used. The amount of data used to evaluate the performance ranges from 10 image pairs to 175 ICA sequences. The performance of the cardiac phase and end-diastolic frame (EDF) detection algorithms were calculated based on an equal weight for all images. Ma et al. (2017) developed two algorithms for contrast inflow detection and the outcomes for the best performing one is shown. As the algorithms are developed for different purposes, the performance cannot be compared.

2.4 Discussion

The aim of this review was to assess the current state of machine learning based algorithms in ICA image analysis. This was done by collecting all relevant literature and reviewing it for: used AI technique and the characteristics of the developed algorithm, including the used image data and model performance. The found literature can be divided into three groups: *segmentation*, *classification* and *other applications.*

Segmentation algorithms

The segmentation algorithms aim to divide ICA images into coronary arteries and background. Most studies used a deep learning approach and specifically a CNN. As the purpose of the segmentation algorithms is to delineate the coronary arteries, the annotations only consist of manual segmentations. Besides ICA images, few other types of data are used in the development of the algorithms. No data from other modalities was used within the segmentation algorithms.

The method for algorithm performance assessment differed between the studies, however in most cases a test-set-approach was used. In addition, it should be noted that the size of the dataset used for evaluation had quite a large range and the metrics used for evaluation varied between the studies. These differences should be considered when comparing the different algorithms, as this makes a direct comparison not straightforward. However, generally all computed metrics lie between 0.75 and 0.99, with some outliers. This indicates that most segmentation algorithms have a reasonable to good performance. The differences in performance are not clearly linked to the type of algorithm or the amount of data used. Possibly, differences in the quality of the used data in combination with the pre-processing steps are the cause of this. Furthermore, the type of data used for calculating the metrics can have a significant influence on the outcomes. It is likely that when a validation set is very similar to the training set, high performance metrics will be found.

Classification algorithms

Classification algorithms can divide images, or parts of images into different categories, based on local features of that image. This is applied in various ways to ICA images in the literature, such as stenosis detection, stenosis classification (lesion type, morphology, significance) and predicting patient related parameters. In most studies a deep learning algorithm, in the form of a CNN, was developed.

In all studies, ICA images were used for training. Besides this, several studies used additional data. This includes FFR assessment, age, BSA, the notion of the involved segment, CCTA images and IVUS data. Data annotation is generally more extensive for classification algorithms as compared to segmentation, as additional information might be needed to divide images into distinct groups. Most algorithms in this group also have a segmentation module. The result of the segmentation module, ICA images with delineated coronary arteries, are then used as input for the classification module. So, a segmentation algorithm can be seen as a tool for image annotation. Besides this, several studies also annotate each vessel with the corresponding anatomical term and other vessel characteristics dependent of the goal of the algorithm.

Again, the studies differ in terms of performance evaluation strategy. However, in most studies test-set-approach was used. As mentioned before, the goals of the classification algorithms vary. Several algorithms were developed solely for the detection of a stenosis, others for classifying the stenosis as significant or non-significant and some algorithms classify the found stenosis into distinct groups based on morphological features. Because these goals are different, the calculated metrics cannot all be compared. Nonetheless, it can be noted that almost every calculated metric lies between 0.70 and 0.99. This indicates that performance generally is reasonable to good. Again, the variability in algorithm performance here is possibly caused by the same factors as in the segmentation algorithms.

Other algorithms

Five of the included studies were neither on segmentation nor on classification. In these studies, several different tasks were performed, some of which might be a useful addition to existing segmentation and classification algorithms. These tasks were catheter detection, cardiac phase detection, image matching, contrast inflow detection and vascular tracking.

Catheter interference is a frequent cause of artefacts in automated ICA image analysis. (73) An algorithm that can correctly identify the catheter in each image can be helpful in reducing these artefacts. The cardiac phase detection algorithm can be used to preselect images for further analysis, in such a way that frames where the arteries are best visible can be selected and calculation on flow can be made. (75) The Image matching algorithm can be used to follow a periodic motion and to generate 3D reconstructions of the heart. (76)) A contrast inflow detection algorithm can be useful for frame selection for further analysis of ICA sequences. (77) A vascular tracking algorithm can track the position of coronary arteries over time. This technique can be used to analyse continuity over time, create a 3D reconstruction over time and assess dynamic information. (78)

Most algorithms are based on deep learning techniques, specifically CNNs. The used data for training depends on the goal of the developed algorithm. To be able to detect a catheter in an ICA image, ICA images with the catheter annotated are needed. For cardiac phase detection ICA images and ECG recordings are coupled. For a full description of all data annotations see the results section and table 2.4. Although the different studies cannot be compared due to the different aim of the algorithms, it is shown in table 2.7 that all metrics, with one exception, lie between 0.87 and 0.99, indicating good

performing classifiers. In turn, this demonstrates the potential value of the algorithms in addition to segmentation and classifications.

Approach to data analysis

Most algorithms developed for automated ICA image analysis are based on deep learning techniques. Only two authors did not evaluate a deep learning algorithm. The most used deep learning approach for the researched tasks in this review was a CNN. It is also, more generally, the most used approach for image analysis. (85) As mentioned before, deep learning outperforms most other machine learning approaches in the field of image analysis. Therefore, the choice for a deep learning approach is expected. However, the performance of the non-deep learning algorithms is not clearly different from the performance of the deep learning algorithms in the reviewed literature. In fact, Hae et al. (2018) evaluated deep learning and other types of algorithms and the best performing one was not deep learning based, all be it with minimal differences. An advantage to deep learning over other types of algorithms is that there is no need for feature engineering. In this way a time-consuming activity can be surpassed. Furthermore, by not predefining features for classification, new relations can be uncovered. On the other hand, there are several disadvantages of deep learning approaches as compared to other types of machine learning. First, after an algorithm is constructed, it is not always clear on what features the classification decisions are based. This non-transparency, obscures choices within clinical decision making and makes it difficult to adequately inform patients. Similarly, non-transparency also introduces the risk of biases in the used data, originating from e. g. human prejudices and collection artifacts, which are propagated into the algorithm. This can lead to decision making, by the algorithm, based on spurious correlations. (86, 87) Second, it is assumed that very large amounts of data are needed for the training of a robust and generalizable deep-learning model, which is very computationally intensive. (70) It might also be problematic to collect a sufficient amount of data for training. However, in most of the reviewed articles, authors have shown that with limited usage of data a fairly well performing algorithms can be constructed. There is quite a large range in the amount of data used for the training of the algorithm. However, multiple studies use data augmentation techniques to increase the actual amount of data used for training. This is done by operations such as rotation or mirroring of the original images. Although it is the general assumption that very large databases are needed, it is shown that an adequately performing algorithm can be developed with limited data.

A great advantage of a deep learning approach to data analysis is that multiple data sources can be easily integrated within the same algorithm. In several studies other types of data were used besides ICA images, such as CCTA images, FFR measurements and patient characteristics. (88, 89) By coupling different data sources, previously unknown relationships or new subpopulations can possibly be identified. In this way AI can really impact clinical decision making and therefore patient outcomes. This can possibly improve healthcare in a more direct way, rather than gaining a slight improvement in segmentation techniques with unknown impact on patient outcomes.

Algorithm performance

All authors evaluated the developed algorithm on a predefined dataset. However, the type of dataset used for evaluation varies. This is important as it influences the performance of the algorithms. As mentioned before, there is generally a choice between a validation set, testing set and an external dataset. Besides this, k-fold cross-validation is also a possibility for performance evaluation. All approaches, in principle, use an unbiased set for evaluation. However, as the validation set is used for parameters fine tuning, the classifier becomes more fitted on the data in the validation set. Calculating performance on this set can give an overestimation of the classifier's performance. To some extend

this also applies to certain testing sets, as it is not uncommon to split one dataset into training, validation and testing. As the training and testing set are very similar, performance can be good in these sets, but much lower when an external dataset is used. Therefore, calculating performance metrics on an external dataset gives the best representation of the algorithm's performance and generalizability. In k-fold cross-validation the dataset is split into k groups. In this approach the performance is calculated k times, with leaving 1 set out of the training set and calculating the performance on this set. The result is the average value over all partitions. In this way the performance is calculated on unseen data and by leaving a part of the data out of the training set, the influence on performance of the left-out data can also be assessed. When leaving one set out, without a strong decrease in performance, this can show that the algorithm is not overfitted on the used training data. (90) The use of different types of datasets, different metrics and the different goals of the algorithms makes comparing them not straightforward. The goal of the segmentation algorithms is similar, they are all aimed at dividing the images into vessels and background. In the classification group the goal is more variable: As mentioned above, some algorithms try to detect stenoses, some try to classify stenoses into different groups of severity and others try to classify the morphology of the detected stenoses. The algorithms that are not in the segmentation or the classification group are even more different in their goal. All this should be part of the assessment of performance metrics.

The performance of most algorithms lies between 0.75-0.99 for segmentation, 0.70-0.99 for classification and 0.87-0.99 for the other algorithms, there are some outliers. This implies that the algorithms perform better than a random classifier, which would typically classify 50% correctly and 50% incorrectly, but it is also not perfect. There are several algorithms with performance metrics below 0.75. When comparing these to the algorithms with the highest performance, no evident differences are present in the amount of data used or in the type of algorithm. This might indicate that differences in performance come from different pre-processing steps or that the quality of the data used is different.

An important question on the performance is how this compares to current methods and approaches. In several studies the performance of other algorithms or other approaches (manual and QCA) are compared to the developed algorithm. It should be noted that in these studies there is no research done on what methods are the current gold standard for the specific procedure. So, it is hard to couple these comparisons to hard conclusions on superior performance.

Several tasks carried out by the developed algorithms can also be performed using classical algorithms (not AI-based). In a study by Cruz-Aceves et al. (2016) a vessel detection and segmentation algorithm was developed. The detection was based on applying multiscale Gabor filters. For segmentation, a multi-objective thresholding approach was used. The algorithm, consisting of both modules, reached a sensitivity of 0.869, a specificity of 0.934, an accuracy of 0.929 and a PPV of 0.802. (91) Qin et al. (2019) also developed a segmentation algorithm based on robust principal component analysis, filtering and tensor completion algorithm. This algorithm reached a sensitivity of 0.773, a PPV of 0.704 and a DSC of 0.729. (92) In a meta-analysis by Collet et al. (2018) the diagnostic performance of image-based computation of FFR was determined. Here a pooled sensitivity of 0.89 (95% CI: 0.83-0.94), specificity of 0.90 (95% CI 0.88-0.92 and an AUC 0.84 (95% CI 0.66-0.94) was found. In a study by Cho et al. (2018) FFR values are also predicted based on ICA images. However, in this study a machine learning approach was used. This algorithm had a sensitivity of 0.80, specificity of 0.87 and an AUC of 0.87. These studies demonstrate that a machine learning approach can be outperformed by a classical algorithm. However, as mentioned before, machine learning algorithms have advantages over classical

algorithms in some areas. Hence, it should be thoroughly investigated per subject what approach leads to the best results.

Limitations

This review explicitly focused on AI algorithms for the analysis of ICA images. This is a limited scope, as there has also been a large amount of research on other forms of data, like electronic health records, laboratory data and text reports, to detect or predict various coronary artery related parameters. Techniques that can play a role in this are data mining and natural language processing. (93) As mentioned before, when combining different sources of information new relations and populations can potentially be discovered. (69) This is especially of interest for diagnosing patients with cardiac ischemia and non-obstructed coronary arteries on an ICA.

Another limitation of this study is the lack of a quantitative comparison between the studies reviewed. Since the studies differed too much, even after dividing the studies into three groups (*segmentation, classification, other*), it is not possible to conclude which algorithm performed best. Within the group of studies on segmentation, the goal of the algorithms is the same, however the way of evaluation differed, as well as the types of metrics. For the studies on classification and the studies on other purposes, not only the way of evaluation and the type of metrics but also the goals of the different algorithms varied.

Relevance for clinical practice

In current clinical practice ICA is used in specific situations. It is used for the diagnosis of CAD when there is a high clinical likelihood of significant CAD. This likelihood is based on initial clinical evaluation (complaints, history, risk-factors). ICA is also an option in patients with suspected CAD and inconclusive non-invasive testing or when non-invasive testing suggests high event risk, so that revascularization options can be determined. Unresponsiveness to medication can be an alternative reason to refer a patient for ICA. (22) When stenoses are graded 40-80%, it is not always clear if these stenoses are hemodynamically significant. In these so-called intermediate severity stenoses, obtaining extra physiological information can give a more conclusive answer. This is most often done via FFR or instantaneous wave-free ratio (iFR). (94, 95)

The reviewed studies differ in clinical relevance. Segmentation algorithms are possibly the least relevant. Currently, automated segmentation of ICA images is not part of the routine. As opposed to other imaging modalities, such as CCTA or CMR, acquisition and analysis are performed concurrently in ICA. In the case of non-concurrent acquisition and analysis, automatic image analysis may be useful, as no professional for reviewing images is needed. In contrast, the cardiologist acquires the images and directly interprets them for decision making. Thus, in diagnostic ICA, automated segmentation may not lead to a significant reduction in procedural time. As mentioned before, the difference in performance between AI-based and conventional segmentation algorithms is minimal. Furthermore, the segmentation algorithms are almost exclusively trained on manually segmented images. This means that a 100% accurate algorithm would perform as good as manual segmentation by an expert. It should, however, be noted that AI-based segmentation can improve the whole process, as it does make segmentation more objective and if multiple segmentations from different experts are used the average value could possibly reduce bias. In addition, automated segmentation is an essential component of reliable navigation software. (96)

As the goals of the classification algorithms differ, so does their relevance. Several classification algorithms focus on detecting stenoses. Since most algorithms are trained on manual annotations,

again, optimal performance would be equal to that of the expert that annotated the images. The relevance of this type of classification algorithms could come from a time saving point of view, as automated stenosis recognition can be done by the algorithm. Algorithms that can predict lesion significance based on ICA images can possibly be of great relevance for clinical practice. It is known that it is not always straight forward to distinguish a significant stenosis from an insignificant one. (37) So an objective and well performing algorithm could be of great help. More specifically, an algorithm like the one developed by Cho et al. (2019), that can predict FFR from the ICA image, could strongly reduce the need for FFR measurement. This can reduce costs, save time and prevent any complications of the FFR measurement. Moreover, the algorithm developed by Cho et al. (2019) outperforms QCA based methods for lesion significance prediction. The algorithms developed by Cho et al. (2019) and Hae et al. (2018) are also good examples of combining data from different sources to classify stenoses. These kinds of algorithms could be of interest in patients with cardiac ischemia with non-obstructed coronary arteries.

The catheter detection, contrast inflow detection and the cardiac phase detection algorithms are also of relevance. These algorithms can be incorporated into other algorithms, as they can preselect images for further analysis and to reduce artefacts caused by the catheter. The image matching and vascular tracking algorithms can be applied to create 3D reconstructions of the coronary arteries. These reconstructions can also possibly be used for further analysis of stenoses, as a 3D model gives more information on the extend of a stenosis than 2D images. However, it should be noted that there are also non-AI algorithms available that can create similar reconstructions with good results. (97)

Future research

In clinical practice there is a discrepancy between coronary anatomy, results of non-invasive testing and the presence of anginal symptoms. (98) There is an increasing number of syndromes without significant stenosis on ICA imaging, but with anginal complaints. This can be as high as 50% of patients undergoing ICA, depending on the studied population and is a frequently described phenomenon in woman with angina. (41, 99)

This is alarming, as it is known that patient with angina and non-obstructed coronary arteries have an increased risk of adverse clinical events. (42) A possible explanation for this discrepancy is that the causes of this condition are not visible on standard ICA or that there are signs present in the images which are not being recognized. This is where AI-based algorithms can be of help. By uncovering yet unknown signs of disease in images of patients with angina without obstructed coronary arteries and by discovering relations between patient parameters which can possibly predict the presence of disease in this group. As is shown in this review, AI analysis of ICA images is a relevant research topic. For further research combining ICA image data with clinical variables would be a valuable next step. There are several possible directions for such research. First, combing ICA images with additional data sources like electronic health records, lab values and other (non)invasive imaging modalities like FFR, IVUS and CCTA. Furthermore, analysis of parameters in the ICA sequences related to flow patterns and temporal analyses could possibly provide more information on abnormalities in the before mentioned patient group. In this way novel associations between different patient characteristics might be uncovered and new (sub)populations might be identified. Another direction could be to determine functional significance of stenoses based solely on ICA images. To decrease the need for additional assessments and with that reduce costs and complications.

As mentioned before, there is a variability in algorithm performance, which is possibly linked to the quality of the used data and the pre-processing steps. Because of this, there should also be a sufficient

focus on the data quality and pre-processing steps in future research. Ultimately, AI should be available as a peri-procedural tool that uses all-available patient data to guide clinical decision making.

2.5 Conclusion

ML-based algorithms for ICA image analysis show good performance in segmentation and several types of classification. However, to further improve clinical practice, the development of future algorithms should focus on solving diagnostic issues and improving clinical outcome. By combining data from different sources and imaging modalities possibly new subpopulations and associations between variables can be uncovered. This can be help with assessment of patients with ischemic heart disease with non-obstructed coronary arteries. Furthermore, ML-based algorithms have the potential to be able to determine significance of stenoses solely on ICA images, when trained with inputs from other modalities. This could decrease the need for additional assessment techniques and reduce costs and complications.

2.6 Tables

Table 2.2. Study characteristics for segmentation literature

Author	Year	Technique	Availability	Data use	Data annotation
Yu F et al. **(100)**	2019	GAN	DRIVE dataset online available Rest: ×	DRIVE dataset: 40 fundus images 1092 ICA images	DRIVE dataset: * ICA images: not annotated
Zhao L et al. **(101)**	2018	CNN	×	60 ICA sequences	*
Yang S et al **(81)**	2018	CNN	×	60 ICA sequences	*
Cervantes-Sanchez F et al. **(102)**	2019	ANN	Database with annotations online available	130 ICA images	*
Wang L et al. **(71)**	2020	CNN	×	170 ICA sequences (8835 images)	*
Ma B et al. **(103)**	2019	ANN	Flying chair dataset online available Rest: ×	727 ICA sequences Flying chair dataset	*
Yang S et al. **(80)**	2019	CNN	Database not publicly available	3302 ICA images	○
Karapataki M et al. **(104)**	1997	Hopfield network	×	2 ICA images	*
Hao D et al. **(105)**	2019	CNN	×	60 ICA sequences	*
Yang S et al. **(106)**	2019	CNN	×	5572 ICA images	*
Plourde M et al. **(107)**	2012	SVM	CAVAREV dataset online available Rest: ×	130 ICA images CAVAREV dataset: angiographic phantom images	*
Fan J et al. **(74)**	2018	FCN	×	148 ICA sequences	*
Cruz-Aceves I et al. **(108)**	2018	ANN	×	100 ICA images	*
Jo K et al. **(109)**	2018	CNN	×	1987 ICA images	○
Nasr-Esfahani E et al. **(110)**	2018	CNN	×	44 ICA images	*
Hao D et al. **(96)**	2020	ANN	Database and source code online available	120 ICA sequences	*
Jun TJ et al. **(111)**	2020	CNN	×	4700 ICA images	○
Nasr-esfahani E et al. **(112)**	2016	CNN	×	44 ICA images	*

GAN: Generative adversarial network, CNN: Convolutional neural network, FCN: Fully convolution network, ANN: Artificial neural network
×: not mentioned, *: manual segmentation, ○: semi-automatic segmentation with manual correction

Table 2.3 Study characteristics for classification literature

Author	Year	Techniques	Availability	Data used	Data annotation
Au B et al. **(72)**	2018	CNN	Data availability: ×, algorithm training details available	1024 ICA RCA images	Manual vessel segmentation, lesion coordinates, QCA measurements of stenosis, lesion silhouette
Cho H et al. **(89)**	2019	Gradient boosting with manual features selection	Not publicly available	ICA sequences, FFR assessment, age, BSA, involved segment of 1501 patients	QCA with manual correction for vessel segmentation, centreline, lumen diameter
Kumar S et al. **(113)**	2014	ANN with predefined extraction	×	100 ICA images of 100 patients	Vessel segmentation with auto-thresholding on enhanced images
Cong C et al. **(114)**	2019	CNN	×	194 ICA sequences	Manual artery labelling and stenosis class labelling and location
Du T et al. **(115)**	2020	cGAN: segmentation CNN: morphology detection	Training details available, rest not mentioned	20612 ICA images	13373 ICA images labelled with coronary artery segments; 7239 images labelled for lesion morphology
Wu W et al. **(116)**	2020	CNN	×	148 ICA sequences	Manual segmentation, bounding boxes on stenoses, stenosis significance labelling
Lessage X et al. **(117)**	2020	CNN	×	675000 ICA images	Labelling arteries: RCA, LCx or LAD and diseased or non-diseased
Chen S et al. **(118)**	2020	CNN	×	14509 ICA sequences	Stenosis labelling: CTO, thrombus or calcification Guide wire labelling
Lee PC et al **(119)**	2019	CNN	Web crawl data, all publicly available data	4980 ICA images	Artery labelling: RCA, LAD LCx, left main Stenosis labelling: clinically significant or not
Liu X et al **(83)**	2019	CNN	×	2059 ICA sequences	Labelling lesions as blunt or tapered morphology
Zhang D et al. **(120)**	2019	CNN	×	105 ICA sequences, 10 frames per viewpoint, 2 viewpoints per patient	Manual stenosis annotation: RVD, MLD, LL1, LL2
Yang S et al. **(84)**	2019	CNN	×	7197 ICA images	Artery labelling, lumen area of major vessel at end-diastolic phase
Hae H et al **(88)**	2018	K-nearest neighbour, binary class L2 penalizes logistic regression, SVM, random forest, extra tree, AdaBoost, light GBM, CatBoost, Gaussian naïve, Bayes, ANN	Database, source code and computed features online available	ICA sequences, FFR assessment and CCTA images from 1143 patients. IVUS for 630 patients	CCTA: centreline of each coronary artery ICA: diameter of stenosis, MLD, lesion length, RVD
Du T et al. **(121)**	2018	CNN	×	100 ICA sequences	Bounding boxes on stenoses
Liu X et al **(82)**	2019	CNN	×	12106 ICA images	Stenosis labelling, bifurcation locations

BSA: body surface area, RVD: reference vessel diameter, MLD: minimum lumen diameter, LL1-2: lesion length1-2

×: not mentioned

Table 2.4. Study characteristics for other literature

Author	Year	Application	Technique	Availability	Data used	Data annotation
Hernandez-Vela A et al. **(73)**	2012	Catheter detection	Adaboost with multiscale stacked sequential learning	Dataset online available	91 ICA images	Manual catheter annotation
Ciusdel C et al. **(75)**	2020	cardiac phase detection	CNN	Not commercially available	45700 ICA sequences with ECG recording	ECG annotated for switch diastolic – systolic phase by automated signal processing, subset manually by expert for evaluation
Fan J et al. **(76)**	2019	Image matching	CNN	KITTI optical flow dataset online available, rest not mentioned	KITTI optical flow dataset (3000 image pairs) and 50 ICA image pairs	KITTI: matching points present in dataset ICA images: manually labelled with 100 correspondence points per image
Ma H et al. **(77)**	2017	Contrast inflow detection	CNN and LSTM-RNN	Not mentioned	120 ICA sequences	Manually marked beginning of contrast inflow for each sequence
Fang H et al. **(78)**	2018	Vascular tracking	CNN	Not mentioned	Single branch dataset: 12 ICA sequences Vessel tree dataset: 9 ICA sequences	Manual delineation of centrelines by 3 experts, average used

LSTM-RNN: long short-term memory recurrent neural network

Table 2.5 Segmentation literature performance

Author	Calculated on	Dataset size	Evaluated algorithm	Accuracy	Precision	DSC	Sensitivity	Specificity	IOU	Error rate	Border error	NPV
Yu F et al. **(100)**	Validation set	328 images	*	0.953 ±0.009	0.820 ±0.031	0.824 ±0.026	0.829 ±0.039					
Zhao L et al. **(101)**	Test set	10 images	*			0.88	0.82	0.96				
Yang S et al. **(81)**	test set	10 images	*		0.8303	0.8007	0.7774	0.9934				
Cervantes-Sanchez F et al. **(102)**	Test set	30 images	*	0.9698	0.7434	0.6857	0.6364	0.9880				
Wang L et al. **(71)**	Test set	1473 images	*				0.779	0.9941	0.818			
Ma B et al. **(103)**	Test set	145 sequences	*			0.72						
Yang S et al. **(80)**	External test set	181 sequences	DenseNet121, whole vascular tree		0.904 ±0.126	0.896 ±0.138	0.898 ±0.155					
Karapataki M et al. **(104)**	External test set	1 image	*									
Hao D et al. **(105)**	Test set	18 sequences	*		0.8604 ±0.0438	0.8437 ±0.0654	0.8319 ±0.0560					
Yang S et al. **(106)**	Test set	1439 images	Densenet121		0.913 ±0.088	0.919 ±0.087	0.926 ±0.096					
Plourde M et al. **(107)**	Test set	49 images	*							2.23%		
Fan J et al. **(74)**	Test set	21 images	*	0.9881	0.8678	0.8725	0.8773	0.9954				
Cruz-Aceves I et al. **(108)**	Test set	50 images	*	0.9568								
Jo K et al. **(109)**	Test set	1788 images	*	0.984	0.764	0.659	0.601	0.995				
Nasr-Esfahani E et al. **(110)**	Test	11 images	*	0.9793		0.8151	0.8676	0.9859			0.3970	
Hao D et al. **(96)**	Test set	30 sequences	3D + CAB algorithm		0.8492 ±0.0605	0.8428 ±0.0531	0.8424 ±0.0813					
Jun TJ et al. **(111)**	Test set	940 images	Optimized T-net, whole vascular tree		0.9050	0.890	0.8832					
Nasr-esfahani E et al. **(112)**	Test set	18 images	*	0.935	0.967		0.90	0.97				0.960

*: end-to-end proposed algorithm

Table 2.6 Classification literature performance

Author	Calculated on	Dataset	Evaluated algorithm	AUC	FDR	A. b.	Sensitivity	Specificity	PPV	NPV	Acc.	F1	MAE	Pearson
Au B et al. **(72)**	Test set	154 images	*	0.703 ±0.033	36.8% ±2.1	-0.1% ±17%								
Cho H et al. **(89)**	Externa test set	79 sequences	*	0.87			0.80	0.87	0.74	0.90	0.85			
Kumar S et al. **(113)**	Test set	50 images	*				0.91	0.88	0.94	0.83	0.89			
Cong C et al. **(114)**	4-fold cross-validation	4-fold cross-validation	Binary classification algorithm								0.85 ±0.02	0.80 ±0.05		
Du T et al. **(115)**	Test set	1031 images	Segment recognition				0.852	0.991	0.762	0.995	0.984			
Du T et al.**(115)**	Test set	1031 images	Lesion morphology classification				0.897		0.762			0.824		
Wu W et al. **(116)**	Test set	25 sequences	Stenosis detection				0.872		0.795			0.832		
Lessage X et al. **(117)**	Validation set	135000 images	*								0.92			
Chen S et al. **(118)**	Test set	Not mentioned	*, multiclass				0.944		0.949			0,946		
Lee PC et al. **(119)**	Test set	149 images	Anatomy classification								0.83			
Lee PC et al **(119)**	Test set	149 images	Stenosis detection								0.74			
Liu X et al **(83)**	Test set	144 sequences	Detection				0.89		0.88			0.89		
Liu X et al. **(83)**	Test set	144 sequences	Morphology classification	0.98			0.95	0.89	0.95	0.89	0.91			
Zhang D et al. **(120)**	10-fold cv	10-fold cv	*										1.27 ±0.71	89.14 ±11.24
Yang S et al. **(84)**	Test set	1439 images	Classification vessel								0.98	0.92		
Hae H et al **(88)**	Externa test set	79 images	*, Catboost	0.89			0.80	0.85	0.71	0.90	0.84			
Du T et al. **(121)**	Test set	10 sequences	*				0.94							
Liu X et al **(82)**	Validation set	3363 images	Bifurcation detection								0.98			
Liu X et al **(82)**	Validation set	3363 images	Lesion detection				0.860	0.829	0.841					

A. b.: Assessment bias, FDR: False discovery rate, Acc.: Accuracy

*: end-to-end proposed algorithm

Table 2.7 Other literature performance

Author	Calculated on	Dataset	Evaluated algorithm	Accuracy	sensitivity	specificity	PPV	NPV	F1	RMS	MAE
Hernandez-Vela A et al. **(73)**	Validation set	20 ICA images	Catheter detection	0.9456	0.5191	0.9863	0.9555				
Ciusdel C et al. **(75)**	External test set	175 ICA sequences	Cardiac phase detection	0.989	0.995	0.981	0.988	0.992			
Ciusdel C et al. **(75)**	External test set	175 ICA sequences	EDF detection		0.993		0.993		0.993		
Fan J et al. **(76)**	Test set	10 ICA image pairs	Image matching							9.79 ±3.9	
Ma H et al. **(77)**	Validation set	80 ICA sequences	RNN-LSTM algorithm								3.6
Fang H et al. **(78)**	Validation set	Not specified	Complex vascular tracking		0.87±0.05		0.89±0.05		0.88±0.05		

3. Thesis feasibility study: Expectations and perceived barriers by interventional cardiologists on artificial intelligence

3.1 Introduction

Despite the promising results in both radiology and cardiology, clinical implementation of AI is still limited. (62, 64) Besides the challenges related to data and model performance, clinical applicability, lack of knowledge on AI by clinicians and lack of trust by clinicians are major challenges in developing and implementing ML models in clinical practice. (66, 122, 123) To ensure clinical applicability, in both a logistic and a medical point of view, clinicians should take the lead in identifying clinical problems that can benefit for ML, as they generally have the best knowledge on problems in clinical practice and will often be the end users of the application. (122) Huisman et al. (2021) conducted a large international survey to assess the levels of knowledge and attitude towards AI among radiologists. In this study it was found that intermediate and advanced knowledge on AI was associated with a more open and proactive attitude towards AI, while lower levels of expertise were associated with a less positive attitude. (124) These findings indicate that increased knowledge on AI might increase clinical adoption of AI models. Clinicians willing to participate in ML research are crucial for thorough validation of AI models in clinical practice and can provide useful feedback. Hence, it is essential to assess the perception of AI by clinicians, to identify where improvements can be made. To assess the perception of AI and to identify clinical applications for AI models in interventional cardiology, a survey was conducted among interventional cardiologists from several hospitals.

3.2 Methods

To evaluate the expectations and perceived barriers on AI by interventional cardiologist and to identify possible applications, a questionnaire was created. The survey was carried out as semi-structured interviews and all participants were interviewed by the author of this thesis. The questions were based on the survey created by Huisman et al. (2021) and adapted based on the objective of the survey and the intended participants. The participants were asked about their level of expertise in AI, the future of interventional cardiology regarding AI, willingness to work with AI, challenges in implementation and possible applications of AI in clinical practice. To identify possible clinical applications and assess the potential perceived by interventional cardiologists, a list of subjects was created. The subjects are shown in table 3.1. The subjects were identified by the author of this thesis together with an interventional cardiologist that did not participate in the survey. The participants were asked to rate the potential of AI models for improving diagnostics and treatment per subject on a scale from 1 to 5. Diagnostics are defined as any application before stenting and treatment as during or after stenting. The subjects were dived in two tables, one for an acute setting and the other for an elective setting. Additionally, the participants were asked if there were any other possible applications that were not presented in the two tables. Furthermore, for each relevant question there was a possibility to add answers that were not listed. Categorical data was described by frequencies (N). Continuous data was described by means ± standard deviation (SD). Answers to open questions were summarized in appropriate tables and figures. All interview questions are shown in Appendix 2.

Table 3.1 Possible clinical applications of AI in interventional cardiology

Acute setting	Elective setting
STEMI MINOCA	Coronary microvascular disease
NSTEMI MINOCA	Vasospastic angina
SCAD	Myocardial bridge
Lesion significance	Generalized diffuse atherosclerosis
Stent choice	Intermediary stenosis
Choice for intravascular imaging	Lesion significance
Complications risk	Stent choice
	Bifurcation stenting approach
	Choice for intravascular imaging
	Predicting restenosis
	Complication risk
	Management of iatrogenic dissections

STEMI: ST-elevation myocardial infarction, NSTEMI: non ST-elevation myocardial infarction, MINOCA: Myocardial infarction without obstructed coronary arteries, SCAD: Spontaneous coronary artery dissection

3.3 Results

A total of 10 interventional cardiologists participated in the survey. Six of the participants work in an academic hospital and four in a general hospital. All participants had some knowledge of AI, six rated their levels of expertise on AI as *"Heard of it"* and four as *"Reasonable knowledge"*.

Future of interventional cardiology and AI

Figure 3.1 shows the response to the question *"Do you think AI has a role in the future of interventional cardiology?"* Six participants thought that AI has a role in the future of cardiology, the other four were not sure about it. No participant responded no to this question.

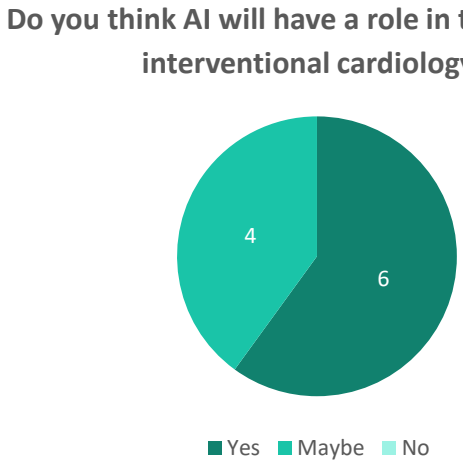


Fig. 3.1 The role of AI in the future of interventional cardiology

In figure 3.2, it is shown what the participants expect to be the influence of AI on interventional cardiology over three time periods. It can be seen that all participants thought that the influence will increase over time, however the extend differs. No participant thought AI will have a very extensive influence at any point in time. All participants thought AI will at least have a minor influence on interventional cardiology within 5 – 10 years.

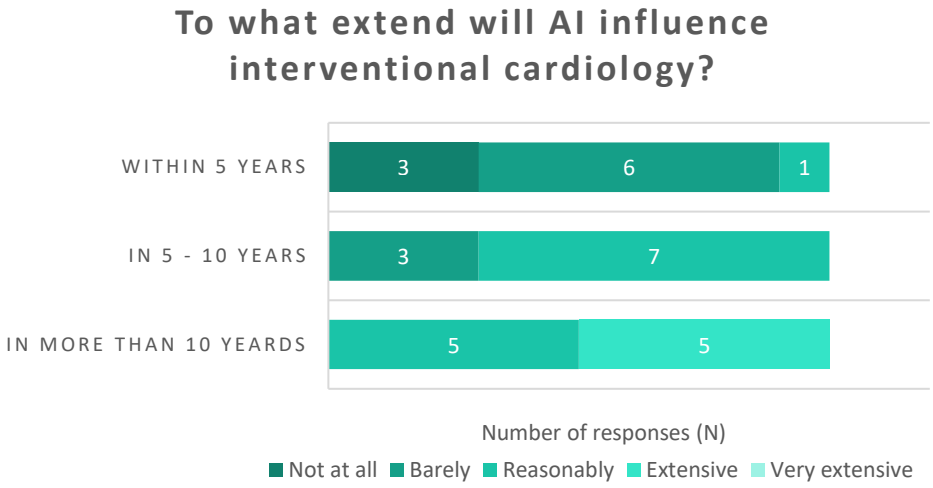


Fig. 3.2 The extent to which AI will influence interventional cardiology on three different terms.

The first effects of AI in clinical practice were expected within different terms. One interventional cardiologist expected the first effects within 2 years, 4 others however thought that it will take at least 10 more years. The average term was 7.2 years, with a SD of 2.9. All participants expected a positive effect of AI on interventional cardiology, where the possible answers ranged from very negative to very positive. The reason for this positive effect were different, as is show in figure 3.3. All participants, except one, expect that AI will play a role in optimizing the workflow, logistic and planning outside the catheterization lab. This includes things like improving selection of patients for ICA, adjusting the daily program based on acute patients and better timing of patient transfer. The number of reasons for a positive effect differs between participants. One participant thought the only reason for a positive effect was the optimization of the workflow, logistics and planning outside the catheterization lab. An average of 3.5 reasons were given for a positive effect, with a SD of 1.4. No participants mentioned effects outside the predefined answers.

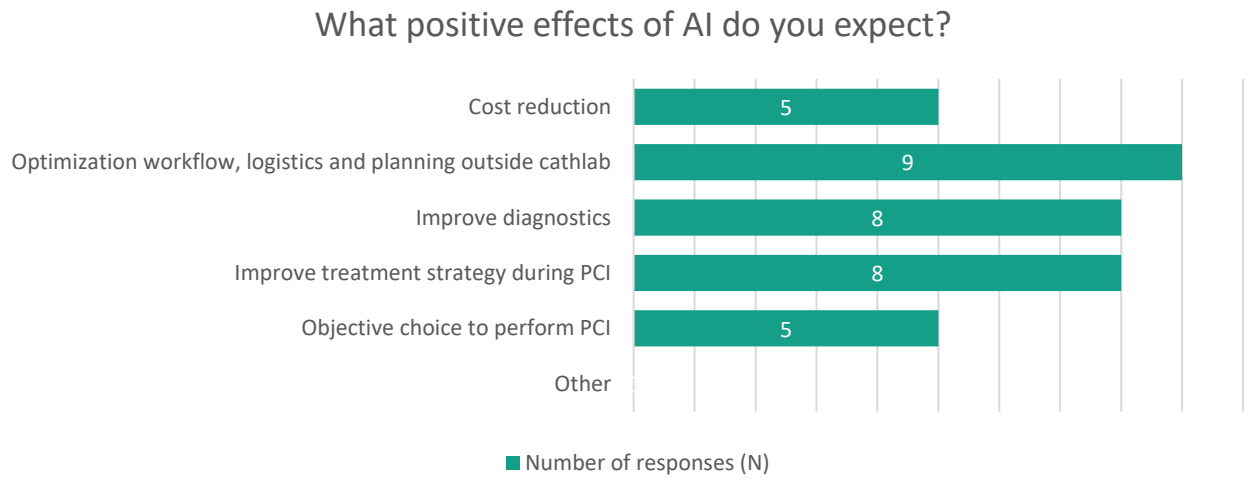
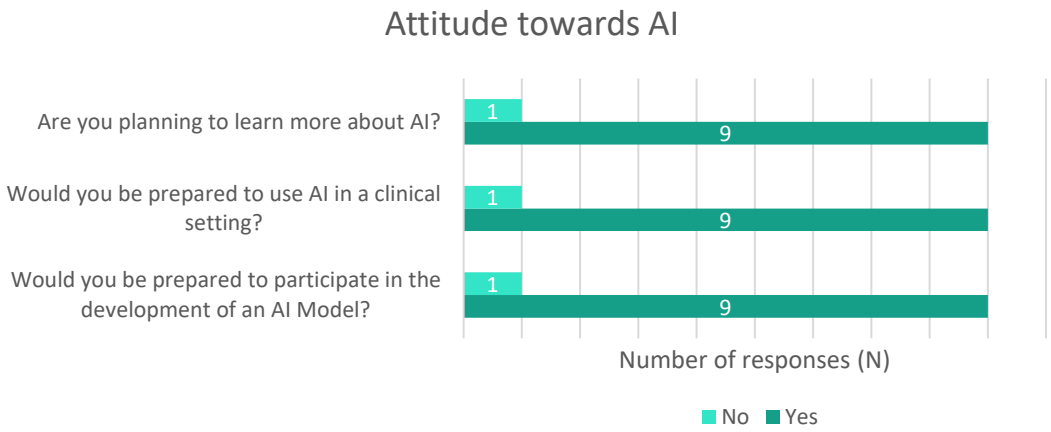


Fig. 3.3 Reasons for the expected positive effect of AI on interventional cardiology.

Attitude towards AI

The attitude towards AI was assessed via three questions: *"Are you planning to learn more about AI?"*, *"Would you be prepared to use AI in a clinical setting?"* and *"Would you be prepared to participate in the development of an AI model?"*. Nine participants answered *"yes"* to all three questions and one participant answered *"no"* to all three. It should be noted that the question regarding the development of an AI model was focused on the clinical applicability and effectiveness and not the development of the actual model.



Fig, 3.4 The attitude towards AI in three questions

Challenges for implementation of AI in clinical practice

The participants were also asked what they think the most relevant challenges for implementing AI models in clinical practice are. They were given 11 possible answers to choose from, including one option *"Other"* which can be chosen if they thought something that was not listed was relevant. They were asked to pick the four most relevant options. The results are shown in figure 3.5 *"Lack of trust in AI by clinicians"* was the most frequently chosen answer (6 times). *"Lack of generalizability"*, *"High costs for software purchase"* and *"High costs for software development"* were all three chosen 5 times. Only one participant proposed an answer that was not listed. The given answer was *"Lack of medical knowledge by ML professionals"*.

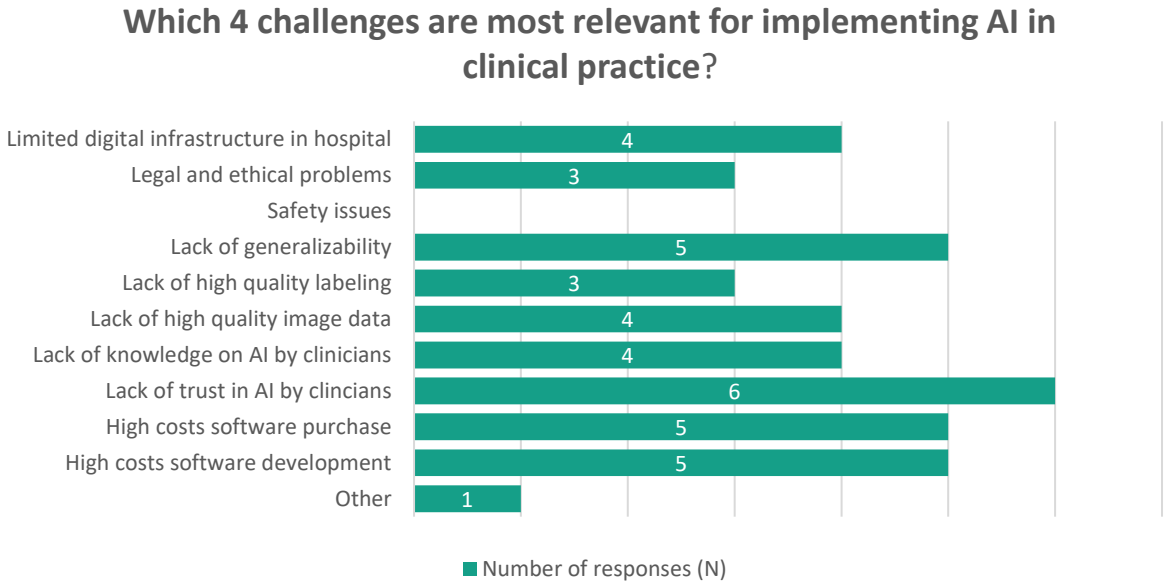


Fig. 3.5 The 4 most relevant challenges for implementing AI in clinical practice.

Clinical applications of AI

To identify possible clinical applications of AI in interventional cardiology, the participants were asked to rate to potential of AI for improving clinical practice for several subjects. The results for the subjects related to an acute setting are shown in figure 3.6-7. The results for the elective setting is shown in figure 3.8-9.

Acute setting

The perceived potential of AI for diagnostics in an acute setting is the highest for SCAD and for therapy in an acute setting for lesion significance. STEMI MINOCA is identified as the subject that benefits the least from AI for diagnostics. Regarding therapy in an acute setting, STEMI and NSTEMI MINOCA are perceived to both benefit the least from an AI model. The average potential value for diagnostics in an acute setting is higher than for therapy (3.1 ±1.4 vs. 2.5 ±1.3), however, the difference is small. The potential of AI is perceived very differently by the different participants, as for several subjects some perceive the potential as very high and other thinks there is no value at all.

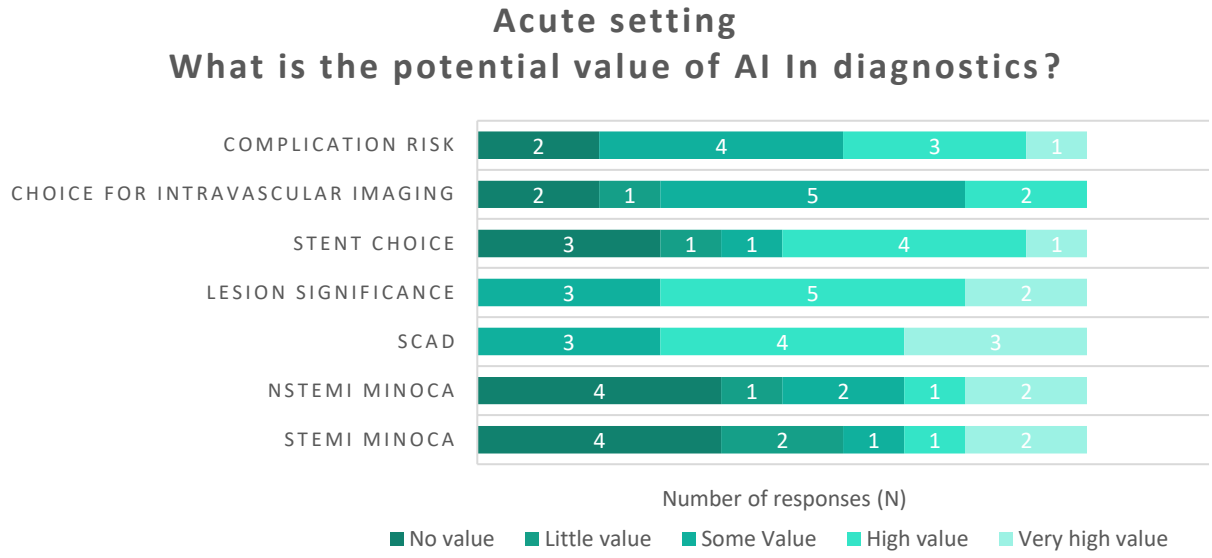


Fig. 3.6 The perceived potential of AI in diagnostics for the acute setting

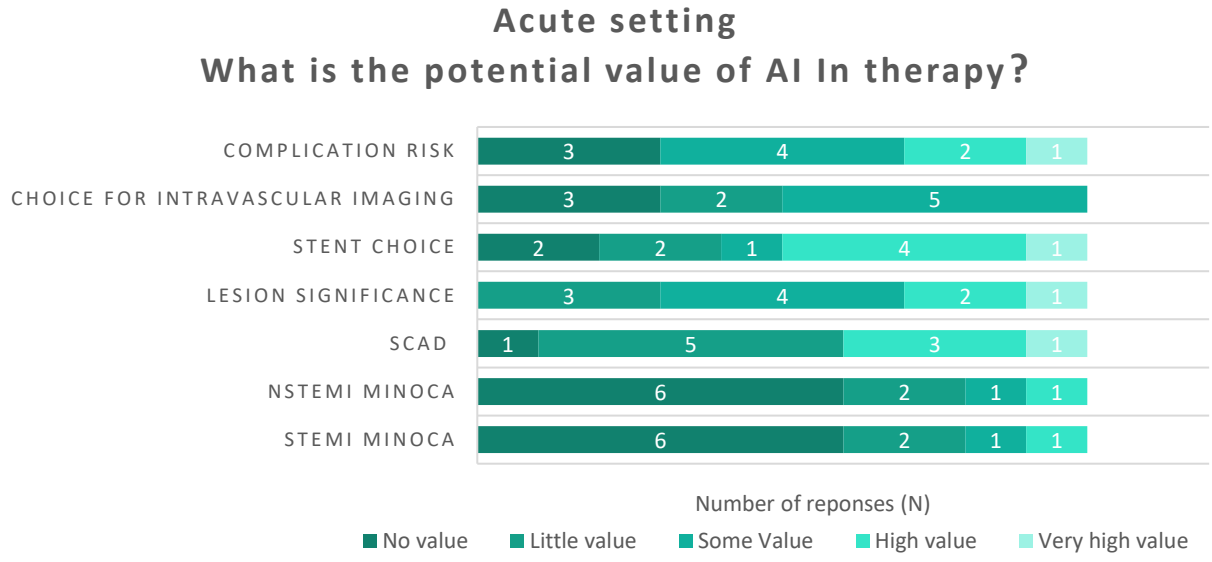


Fig. 3.7 The perceived potential of AI in therapy for the acute setting

Elective setting

The perceived potential of AI for diagnostics and therapy in an elective setting is the highest for lesion significance. The perceived potential for both diagnostics and therapy was lowest for vasospastic angina. The perceived potential of AI for diagnostics an elective setting is higher than for therapy (3.3 ±1.1 vs. 2.8 ±1.2), however the difference is small. In the elective setting there are also large differences between participants. The average potential value for the elective setting was higher than for the acute setting (3.0 ±1.2 vs. 2.8 ±1.3). The difference is, again, small.

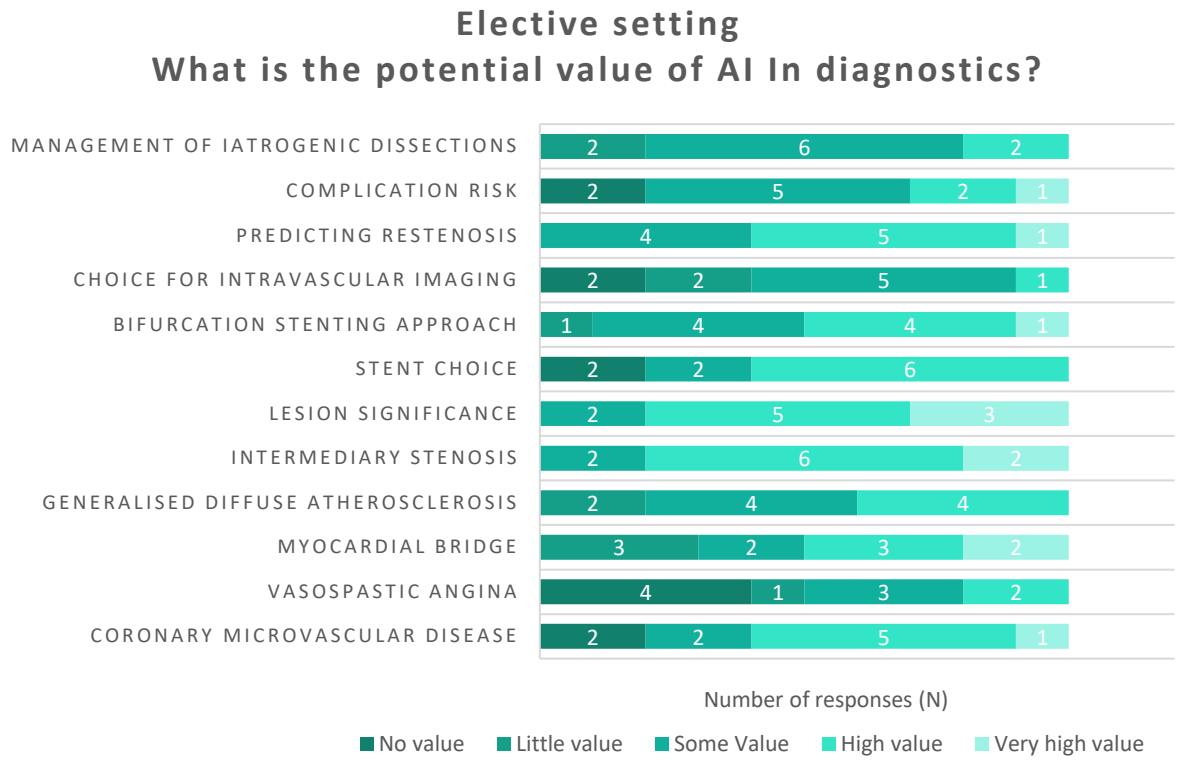


Fig. 3.8 The perceived potential of AI in diagnostics for the elective setting

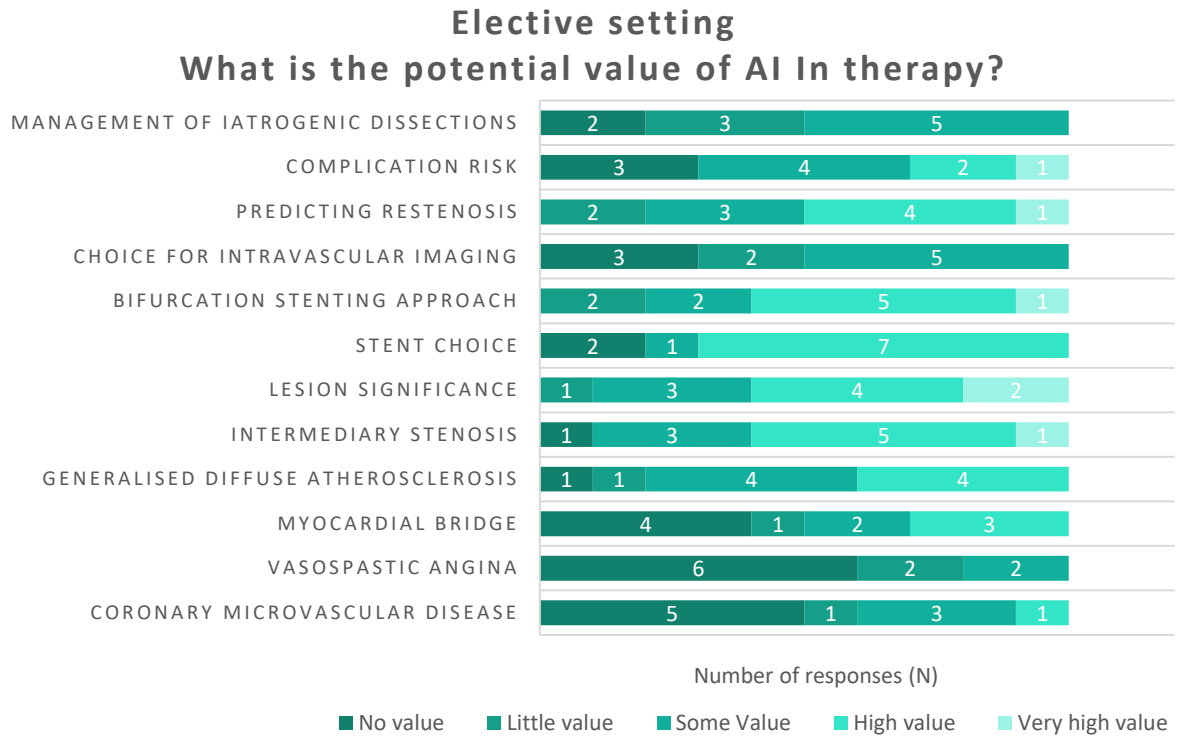


Fig. 3.8 The perceived potential of AI in therapy for the elective setting

Additionally identified applications

Besides the possible clinical applications identified before conducting the survey, several participants mentioned additional possible applications. The mentioned subjects are summarized in table 3.2. Coronary anomalies are a diverse group of congenital disorders. The most common forms are an anomalous location of the coronary ostium and the absence of the left main trunk. (10) Coronary artery ectasia is the dilatation of an arterial segment to at least 1.5 times that of the adjacent normal coronary artery. (125) With debulking techniques like lithoplasty and rotablation are meant.

Table 3.1 Additionally identified clinical applications

Diagnostic	Therapy	Other
Detecting coronary anomalies and determining need for intervention	Predicting debulking results and optimizing results	Translating alarm signs to actions
Assessing relevance of coronary artery ectasia	Determining risk factors for stent complications	Determining day treatment or admission
Predicting coronary occlusion in TAVI procedure		Best access site for sheath
Predicting the need for debulking techniques		
Quantification of calcifications		
Adjustment of patient information brochure based on images		
Patient selection before ICA (other modalities)		

TAVI: Transcatheter Aortic Valve Implantation

3.4 Discussion

To assess the expectations and perceived barriers regarding AI in interventional cardiology and possible clinical applications, a survey of interventional cardiologist was carried out. In total, ten interventional cardiologists were interviewed.

All participants had heard of AI before and four of them rated their level of expertise on AI as reasonable knowledge. However, as this is based on their personal perception it might not always be accurate. During the interviews it became clear that some participants did not fully understood the concepts of AI, specifically ML and DL. It is not necessary for clinicians to have in-depth technical knowledge on AI, however, conceptual knowledge of this subject is necessary for further acceptance and development of clinical AI. (66, 122) Most participants rated the potential value of AI for MINOCA as no or little value. The reason given for this low value was that there are no angiographic features in MINOCA that can be used. This illustrates that the concepts of DL are not entirely understood, as DL is able to find features in data that are not seen by humans or other algorithms. (126)

All participants thought AI will eventually influence interventional cardiology, with an increasing influence over time. To what extend and within what term was perceived differently across participants. One participant expects the first effect in clinical practice within 2 years and mentioned that one could argue that AI is already having effects on clinical practice, but not within the catheterization labs yet. Four participants expect that it will take at least 10 years before there will be any effect. However, all participants think that the influence of AI will be at least reasonable in 10 years. All participants thinks that AI will have a positive influence on interventional cardiology, however the reasons for this positive effect differ. All participants except one think that AI will improve logistics and planning outside the catheterization lab or in preparation for catherization. One participant thought this was the only reason for a positive effect of AI and stated that AI cannot be used in the catherization lab as the imaging data is not fitted for this. Eight out of ten think that AI will improve diagnostics and treatment strategy during PCI. Five out of ten think that AI will reduce costs and that AI will make it possible to make more objective choices for PCI. This shows that generally the interventional cardiologists that participated think that AI can be of value for their field of work. This also becomes clear when looking at the results on the questions regarding attitude towards AI. All but one participant would like to learn more about the subject, are willing to use AI in clinical practice and are willing to collaborate in the development of an AI model. This is a valuable observation, as it is impossible to develop and implement clinically applicable AI models without the collaboration of clinicians. This collaboration can consist of defining clinical questions of interest, guiding choices for optimization of the model and validation of clinical effectiveness.

When asked about challenges for implementing AI in clinical practice, the most chosen answer was lack of trust in AI by clinicians. Lack of generalizability, high costs for software purchase and high cost for software development were all three chosen five times. The participants were less worried about data and labeling quality and safety issues was not chosen at all. This can give an insight in the perception of AI by interventional cardiologists. Lack of trust was most often chosen as a challenge for implementation, while all participants think that AI will have a positive influence on interventional cardiology. This lack of trust can possibly originate from a lack of knowledge on AI, as was also shown by Huisman et al. (2021). Another possible source for the lack of trust is a lack of evidence for clinical effectiveness. A possible way to overcome these issues is to further educate clinicians on AI and by generating data on clinical effectiveness in a safe way. It is also interesting to see that lack of high-quality data and labels was not chosen that much, while most participants mentioned during the interview that the image quality in ICA can be a major problem for AI. It should be noted that these

challenges related to data are also related to the lack of generalizability of AI models. As improper data for training will result in models that are not generalizable. Limitations in digital infrastructure was also chosen four times. Problems with technical integration are known to be a major bottleneck in implementing AI in clinical practice. (122, 127) However, considering the information pipeline and infrastructure requirements thoroughly can identify problems at an early stage and reduce them. (122)

The perceived potential of AI for solving clinical problems varies strongly over the participants. There are several possible causes for this. The extend of knowledge on AI can influence the perceived potential. (124) During the interviews it became clear that personal interest might also be a factor of influence. Several participants also mentioned during the interviews that some subjects were of more interest to them than others. Furthermore, the work environment may also be a factor of influence. One participant, working in a general hospital, mentioned that subjects like MINOCA might be more of interest in academic centers, to gain more insight in pathophysiology of the condition. This participant also mentioned that in a general hospital the focus is generally more on production (performing diagnostic ICA's and PCI's) rather than research. The highest perceived potential for diagnostics in an acute setting was for SCAD. Based on the responses of the participants this potential came from a combination of detectable angiographic features and clinical relevance. The subject with the second highest perceived potential for diagnostics in an acute setting was lesion significance, again based on a combination of detectable angiographic features and clinical relevance. For therapy in an acute setting lesion significance was the subject with the highest perceived potential, followed by SCAD. STEMI and NSTEMI MINOCA had the lowest perceived potential, however most participants assigned a low potential value to these subjects not only on clinical relevance, but also based on the perceived chance of developing a successful model. For diagnostics in an elective setting the perceived potential was highest for lesion significance, followed by intermediary stenoses. For therapy the same two subjects were rated highest. Most participants mentioned that these two subjects are very closely related. Generally, diagnostic applications are rated slightly higher than therapy and the elective setting higher than the acute setting. However, the differences are small.

The results of this survey give a global overview of the perception of AI by interventional cardiologists. However, as only 10 participants were interviewed, the sample size is rather small. For further research it would be interesting to increase the sample size, so that a more representative overview can be created. With a larger sample size it would also be interesting to do a subgroup analysis, to investigate the influence of factors like work environment and years of experience on the perception of AI.

3.5 Conclusion

In this survey it became clear that the level of knowledge on AI by interventional cardiologist is limited. However, generally, the expectations for the future are positive, just as the attitude towards AI and the willingness to collaborate in developing AI models. It should be noted that a very proactive attitude should not be expected, as most participants indicate that their primary focus is not on gaining in-depth knowledge on AI. A good way of involving interventional cardiologists, would be regular scientific presentations on AI, challenging them with questions and involving them in parts of the process that are interesting from a clinical point of view, like defining clinical questions and setting up studies for validation of clinical effectiveness. Based on the perceived potential, developing an AI-based algorithm for determining lesion significance and significance of intermediary stenoses would be a good step to demonstrate the potential of AI in a way closely related to the clinical practice. Furthermore, it would be interesting to investigate subjects like MINOCA, for which the perceived potential is fairly low, to investigate if complex syndromes with no clear angiographic feature could benefit from AI-based models.

4. Querying Electronic health record systems for the creation of ML datasets

4.1 Introduction

Datasets that accurately reflect real-life patient populations with accurate labeling are essential for the development of well performing ML models. (64, 66, 122) Generally, increasing dataset size is associated with higher model performance. (126) However, creating sufficiently large datasets can be challenging. (64) Electronic health records (EHRs) hold massive amounts of data, which can be accessed retrospectively. This has several advantages over prospectively gathering data. Patients do not have to undergo additional procedures, reducing cost and risks for patients. Additionally, EHRs contain longitudinal data regarding patient status, received care and disease progression. This type of data has been shown useful for supporting clinical decision making, diagnosis, risk assessment and medical concept extraction. (128)

EHRs also pose several challenges for the creation of datasets for the development of ML models. Most EHRs were not created for large scale data extraction, but to replace paperwork. Due to this nature, EHRs are often not standardized and based largely on free-text reports. (129) Furthermore, it is not uncommon that hospitals created their own data management system for EHRs, so data entry can be fairly different across institutions. (130)

In the Netherlands, reimbursement for treatment by health insurers is based on Dutch health insurance codes (DBC-codes; Dutch: diagnose-behandel combinatie codes). DBC-codes are six-digit codes describing the healthcare services and prices coupled to it. These codes are used in retrospective research on clinical data. (131-133) Therefore, we hypothesized that DBC-codes can be used as a tool for data labelling within ML datasets. To assess the feasibility of this approach, we systematically tested the yield of queries based on DBC-codes.

4.2 Methods

MINOCA is defined as the presence of an acute myocardial infraction (based on the 4th universal definition of AMI, shown in table 1.2, in Chapter 1), non-obstructed coronary arteries on ICA (defined as no coronary stenosis ≥50% in any potential infract-related artery) and no clinically specific cause for the acute presentation. (59) Myocardial infarctions are divided over two DBC-codes: Non-ST elevated myocardial infarction (320-205) and ST-elevated myocardial infarction (320-204). The choice was made to also include the DBC-code unstable angina pectoris (AP) (320-203), as these three diagnoses are closely related and are managed in very similar ways. In this way MINOCA patients labelled with unstable AP are not missed in this analysis. These three DBC-codes are used as starting points for the creation of the queries.

All queries were created in collaboration with the software engineering team that manages and developed the electronic health records system used in the cardiology department of the LUMC. Since 2006, the department of cardiology and thoracic surgery in the LUMC uses a self-developed electronic patient record system: EPD-Vision. (134) All queries are created in SQL (structured query language). The queries are created based on the definition of MINOCA. The acute myocardial infraction criterium is implemented by using the DBC-codes. To make sure only patients with non-obstructed coronary arteries are included, all patients that underwent coronary artery interventions were excluded. Interventional treatment of CAD is done by PCI or CABG. (22) Per cardiac catherization the carried-out procedures are registered. To exclude all patients that underwent PCI, the condition NO “PTCA” and NO “stenting” were added to the query. To exclude patients that underwent CABG based on the ICA

images, the condition NO “CABG” was added. For MINOCA the coronary arteries have to be proven to be not obstructed on ICA. To ensure this, all patients that did not underwent ICA were excluded. To exclude patients with a clinically specific cause for the acute presentation, such as pulmonary embolism or pericarditis, all patients included by the query were checked manually, by reviewing status reports. Furthermore, to include patients with a possibly incorrect DBC-code, all patients included in the query with DBC-code unstable AP were manually checked for the presence of an AMI. After each procedure the operator (interventional cardiologist) indicates the presence and the extend of CAD as a summary. This is done in a multiple-choice menu. The output of the queries consisted of the patient identification number and the summary for each included patient. The CAD summary might also be of interest for querying.

Additionally, a text-based search was conducted. This search was done in the cardiac catherization reports and the discussion and conclusion part of the consultation letter. The search terms used for this were “MINOCA”, “Gepasseerd thrombus” and “Gepasseerd stolsel”, as these are the most used terms for MINOCA patients.

All queries were carried out on data acquired in 2019 and 2020. For all patients included by the queries, the EHRs were checked for the actual diagnosis. This shows how much of the included patients actually were diagnosed with MINOCA.

Table 4.1 Implementation of the MINOCA criteria for querying

MINOCA criteria	**Implementation**
AMI	DBC-codes: - NSTEMI (320-205) - STEMI (320-204) - Unstable AP (320-203)
Non-obstructed coronary arteries on ICA	Coronary angiography NO PTCA NO stenting NO CABG
No clinically specific cause for acute presentation	Manual analysis of EHR

4.3 Results

Three queries based on DBC-codes and six queries based on text were carried out.

STEMI

The query with the DBC-code STEMI resulted in 18 patients. A flowchart representation of the query results and the different subgroups of patients is shown in figure 4.1. Out of the 18 patients that were found, 7 were CAD patients (with obstructive coronary artery disease). These patients showed up in the query results for several reasons. One patient underwent thrombosuction and PTCA, however the PTCA was not correctly registered. Three patients with significant CAD did not underwent intervention, as the choice was made for conservative treatment. This choice was made based on patient characteristics as age and comorbidities and on the estimated chance of success. One patient underwent ICA in the LUMC and was referred for CABG in another hospital. Another patient with CAD was stented, however this was not correctly registered. One patient with CAD died before any intervention could be done. Seven patients had non-obstructive coronary artery disease (nCAD). Four of these patients were classified as MINOCA, two as takotsubo cardiomyopathy and one patient as chemotherapy induced vasospasms. It should be noted that for the MINOCA and takotsubo patients the classification given was based on the most likely diagnosis. For these patients most EHRs state that the differential diagnosis contains both MINOCA and takotsubo. Four patients had other clinical causes for their presentation. One patient had a ventricular septal rupture, after earlier myocardial infarction. This patient went for surgery after ICA, but not CABG. One patient was diagnosed with SCAD and one with pericarditis, based on CMR. One patient underwent ECMO (extracorporeal membranous oxygenator) placement, which is done under fluoroscopic guidance. The query resulted in 18 patients of which, 7 were in the intended group. This means that approximately 39% of the found patients were part of the intended group.

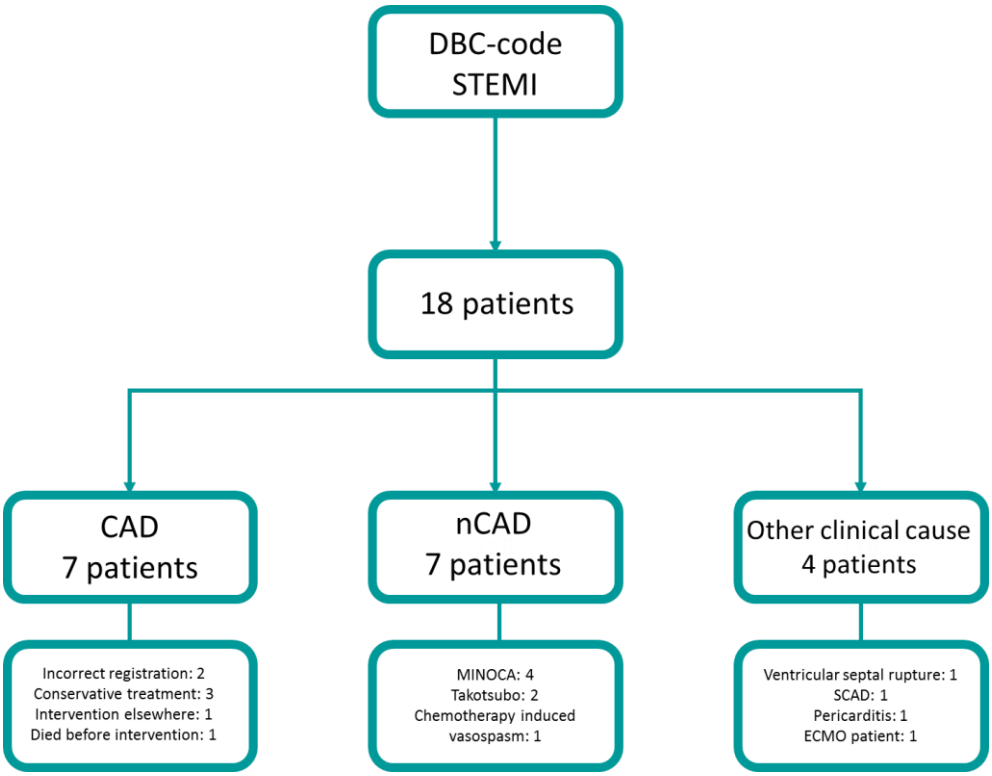


Fig. 4.1 Results of database query based on the STEMI DBC-code.

NSTEMI MINOCA

The query for NSTEMI MINOCA resulted in 50 patients. A flowchart representation of the query results and the different subgroups of patients is shown in figure 4.2. Ten patients in this query were classified as CAD. Out of these 10, five were treated conservatively, four had interventions (PCI and CABG) done elsewhere and one patient died before any intervention could take place. 31 patients were classified as nCAD. The EHRs of these patients stated for 30 patients that MINOCA was the most likely cause and for one Takotsubo cardiomyopathy. However, no definitive results were given. Nine patients had other clinical causes for their presentation. One patient had an atrioventricular block. One patient had a slit-like left main branch, for which an unroofing procedure was carried out. Another patient had a thrombosis of the left main branch. Four patients were diagnosed with SCAD, one with gastro-intestinal problems as a cause of the presentation and one with myocarditis. The query resulted in 50 patients of which, 31 were in the intended group. This means that approximately 62% of the found patients were part of the intended group.

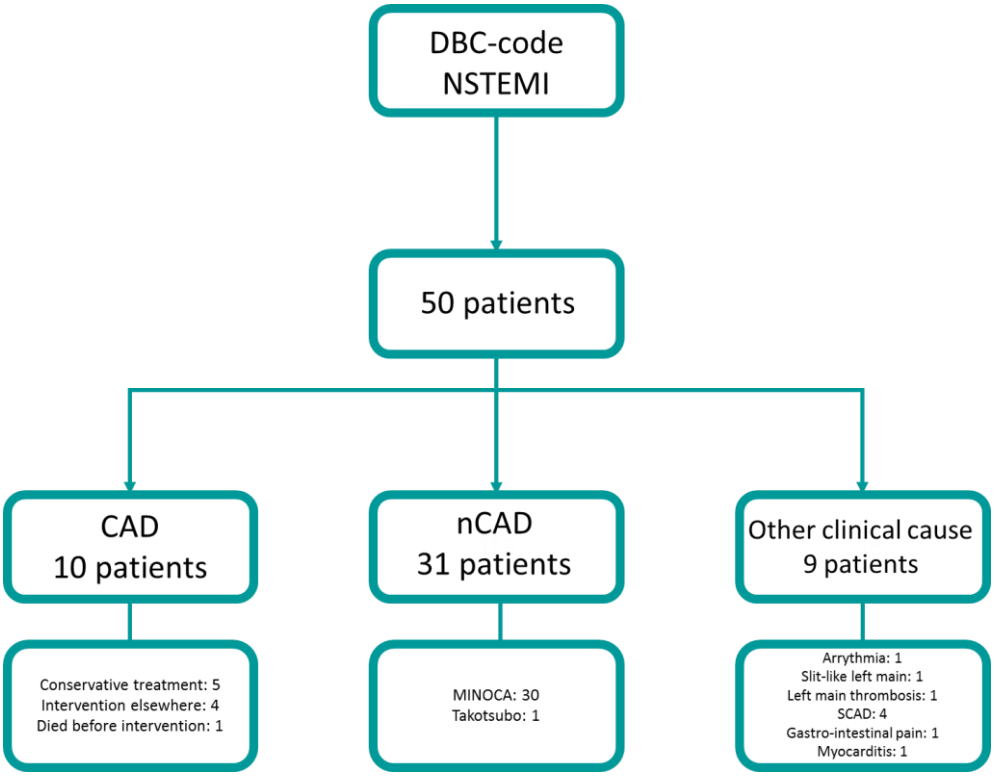


Fig. 4.2 Results of database query based on the NSTEMI DBC-code.

Unstable AP

The query based on the unstable AP DBC-code resulted in 32 patients. A flowchart representation of the query results and the different subgroups of patients is shown in figure 4.3. Six of these patients were classified as CAD. Three of these six patients received conservative treatment, two were transferred for intervention elsewhere and one died before any intervention could take place. A total of 22 patients were classified as nCAD. None of these patients suffered an AMI, so none are classified as MINOCA. 17 patients were classified as INOCA, however, for these patients it was also stated that a non-cardiac cause cannot be excluded. For two patients myocardial bridging was seen as culprit for the presentation, for two other patients vasospasms were seen as culprit and for one patient microvascular disease. Four patients had other clinical causes for their presentation. One patient suffered from arrythmia, one from hypertrophic obstructive cardiomyopathy (HOCM), one from pericarditis and one from an aortic valve stenosis. The query resulted in 32 patients of which 22 were

classified as nCAD. This means that approximately 69% of the found patients were part of the intended group. However, no patients were classified as MINOCA.

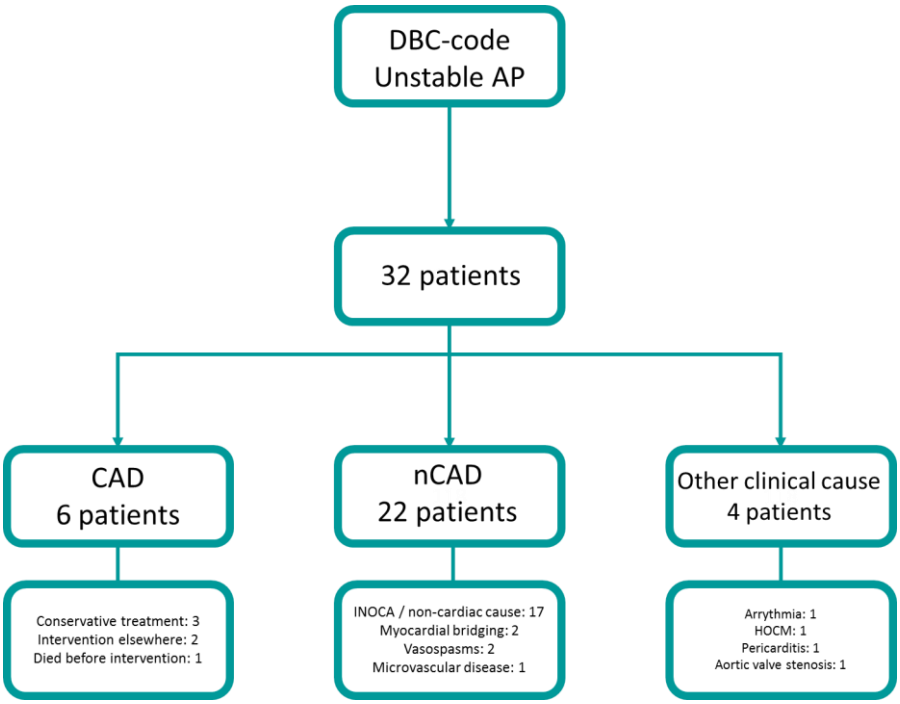


Fig. 4.3 Results of database query based on the unstable AP DBC-code.

Coronary anatomy summary

In the STEMI query for 4/18 (approximately 22%) patients the coronary anatomy summary was not filled in and for one patient classified as MINOCA the summary stated 1 vessel disease. The catheterization report stated not significant CAD. In the NSTEMI query for 15/50 (30%) patients the coronary anatomy summary was not filled in. In one patient classified as MINOCA the summary stated 2 vessel disease, while the catheterization report stated no significant CAD. In the unstable AP query the coronary anatomy summary was not filled in for 11/32 patients (approximately 34%). For one patient classified as myocardial bridging the summary stated 1 vessel disease, while the catheterization report states no significant CAD.

Text-based queries

The text-bases queries were carried out on the catheterization reports and the consultations letters, as described in the methods section.

Consultation letter

The consultations letters that were created in EPDV in 2019-2020 that are now locked (no further adjustments possible) were searched for the three terms commonly used in the context of MINOCA. The letters have to be locked for the SQL query to function properly. The fields “Discussion” and “Conclusion” were searched for the three commonly used terms to describe MINOCA. Flowchart representations of the query results and the different subgroups of patients are shown in figure 4.4.

MINOCA

Querying with “MINOCA” resulted in 57 hits, with 26 patients turning up multiple times in the query. After removal of duplicates 31 patients remained. 14 patients found in this query were also found in the DBC-code based queries, 13 in the NSTEMI query and one in the unstable AP query. The other 17

patients were not found with the other queries. For 12 patients the ICA was carried out before 2019 and so they were not found with the DBC-code. They do show up in the text-based query, as the conclusion MINOCA is repeated in the letters. The other five patients were classified under other DBC-codes: 2 as thoracic complaints e.c.i. (e causa ignota; of unknown origin), one as pericarditis and two as stable AP. For one of the patients with DBC-code thoracic complaints e.c.i. the term MINOCA is used incorrectly. This patient did not have ECG-changes or elevation of troponin, so there was no AMI. Furthermore, in the discussion part of the letter MINOCA is defined as "*Myocardial angina with normal coronary arteries"*, additionally it was stated that vasospasm or microvascular disease possibly causes the complaint. Based on this, the patient should be classified as INOCA. The other patient with DBC-code thoracic complaints did have a NSTEMI (typical acute complaints and rise in cardiac troponin), the letter also *states "Concluding, this patient seems to have a myocardial infarction without obstructed coronary arteries*", based on this the correct DBC-code for this patient is NSTEMI. The letter for patient with DBC-code pericarditis stated that based on the persistent complaints MINOCA is unlikely and that the differential diagnosis should contain peri-myocarditis. After catheterization, this patient was transferred to another hospital for further investigation. Due to this no definitive conclusion is present in EPDV. For one patient with DBC-code stable AP the referral letter from another hospital stated that the patient was suffering a stuttering STEMI (typical complaints, highly increased cardiac troponin and ST-elevations on the ECG that disappeared later on). Based on this, the patient should have been labeled with the DBC-code STEMI or NSTEMI. For the other patient with DBC-code stable AP the referral letter from another hospital stated that the patient was suffering a NSTEMI at the time of transfer.

Passed cloth

The text-based search in the consultation letter for the term "Gepasseerd stolsel" resulted in 28 hits. The results contained 10 duplicates. Eight patients found in this query were also found in the DBC-code based queries, six in the NSTEMI and two in the STEMI query. Out of the 10 newly identified patients, 2 underwent ICA before 2019. Two other patients were filtered out of the DBC queries due to incorrect registration of underwent procedures, PTCA was registered for both, however based on the cathlab technician reports these patients did not underwent PTCA. Both patients were classified as MINOCA. Another patient had a history of CABG, which means that this patient suffers from CAD. This excludes the diagnosis MINOCA. For five patients the DBC-code was different from the ones used in the DBC queries. Two patients were labeled with the DBC-code thoracic complaints e.c.i. For one of these patients the letter stated that the patient suffered a STEMI. The other patient with DBC-code thoracic complaints e.c.i. developed complaints after an earlier ICA and the letter described that this was possibly caused by an air embolism or a passed cloth. Furthermore, this patient had a history of PCI's, hence MINOCA is not possible. One patient was labeled with the DBC-code supraventricular arrythmia. This patient suffered a STEMI in anther hospital, where it was classified as MINOCA. This conclusion was repeated in the discussion part of the letter. Due to this the patient was found in this query. The care this patient received in the LUMC was unrelated to this, which is why this patient was not found in the DBC queries. One of the patients labeled with the DBC-code stable AP was also found in the text-based query on MINOCA. The letter stated: *"differential diagnosis MINOCA, passed cloth"* The letter from the referring hospital stated that this patient suffered a stuttering STEMI. For the other patient with the DBC-code stable AP, the letters stated that the patient was suffering a NSTEMI. Furthermore, this patient underwent PCI before, so MINOCA is excluded.

Passed thrombus

Querying the discussion and conclusion sections of the consultation letters with "Gepasseerd thrombus" resulted in 20 hits, with 6 patients showing up multiple times. Five patients found in this

query were also found in the DBC-code based queries, two in the STEMI query and three in the NSTEMI query. Out of the 9 newly identified patients, 4 underwent ICA before 2019 and in one case "Gepasseerd thrombus" referred to a cerebral infarction. The other four patients were labeled with other DBC-codes. In two cases the DBC-code stable AP was used. One of these patients suffered a NSTEMI in another hospital and was referred to the LUMC for periodic controls. At the time of control in the LUMC the patient was correctly labeled as stable AP. Based on the letters and catherization reports this patient could be classified as INOCA. The other patients labeled with DBC-code stable AP was also correctly labeled with this. However, this patient had a slight rise in troponin levels in another hospital, without significant coronary stenosis. This patient could also possibly be labeled INOCA. The same goes for the patient labeled with thoracic complaints e.c.i. This patient also had a slight rise of troponin levels without significant coronary stenosis on ICA. However, for these three patients there was not sufficient information present to rule out any non-cardiac cause for their presentation. One patient was labeled with the DBC-code endocarditis. This patient had a rise of troponin levels and typical complaints. However, infection parameters were also increased and no significant stenoses were found during ICA. Based on this the diagnosis endocarditis is most likely

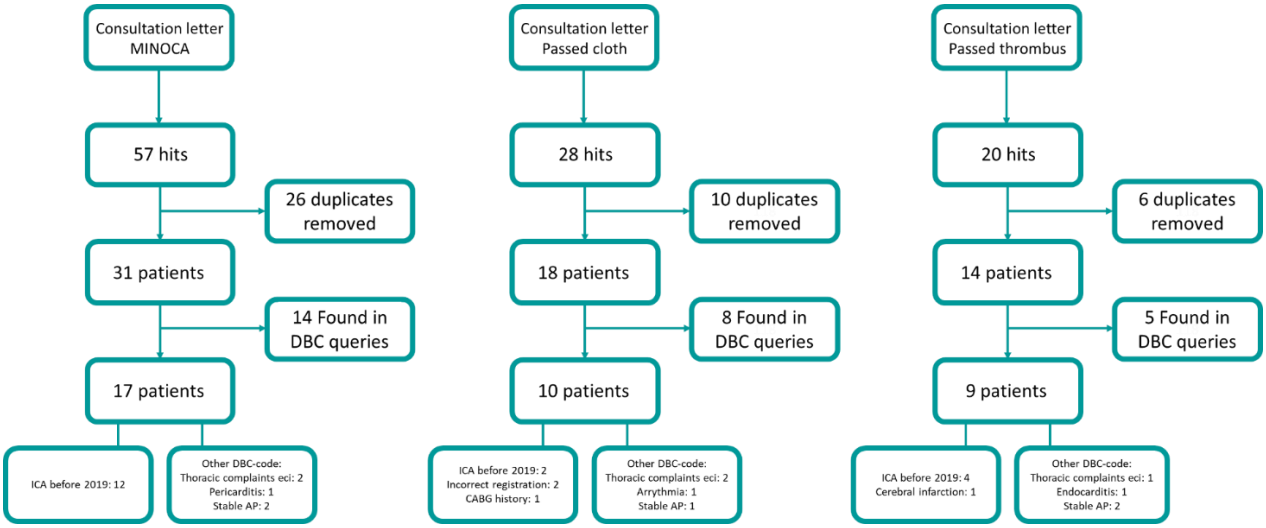


Fig. 4.4 Flowchart for the three text-based queries of the discussion and conclusion section of the consultation letters.

Catheterization report

All catheterization reports created in 2019-2020 were queried for the terms "MINOCA", "Passed cloth" and "Passed thrombus". However, not a single hit was found.

4.4 Discussion

Creating sufficiently large and accurately labeled datasets for ML on medical images is challenging. (64) EHRs hold massive amounts of data regarding patient status, received care and disease progression. However, EHRs are often not standardized and based largely on free-text reports. (129). To access this data in a scalable way a structured approach is needed. The Dutch DBC-coding system for healthcare services, systematically describes diseases and healthcare services. These codes are used in retrospective analysis of clinical data. (131-133) The feasibility of querying EHRs for the creation of ML datasets based on DBC-codes to solve diagnostic challenges in MINOCA patients was assessed by conducting three DBC-codes queries. Additionally, six text-based queries were carried out to investigate if any important DBC-codes were missed and to compare both approaches. The used DBC-codes were NSTEMI, STEMI and unstable AP. The terms used for the text-based queries were "MINOCA", "Gepasseerd stolsel" (passed cloth) and "Gepasseerd thrombus" (passed thrombus). For each of these terms a query was carried out on the consultation letters (discussion and conclusion section) and the catheterization reports.

DBC-queries

The query with DBC-code STEMI resulted in 18 patients, of which 7 were classified as nCAD (39%). The NSTEMI query resulted in 50 patients, with 31 classified as nCAD (62%) and the unstable AP query resulted in 32 patients with 22 classified as nCAD (69%). The patients classified as nCAD in the STEMI and NSTEM query consisted of patients diagnosed with MINOCA (34 cases), Takotsubo (3 cases) and chemotherapy induced vasospasms (1 case). None of the nCAD patients in the unstable AP query suffered an AMI, so no patient was classified as MINOCA. In 17 of these patients the cause of the symptoms was not clear. They were labeled as INOCA, with the addition that a non-cardiac cause cannot be excluded. For 5 patients in the nCAD group from the unstable AP query a culprit was suggested. For two patients this was myocardial bridging, for two other patients this was vasospasms and for one patient this was microvascular disease. However, definitive proof was not always found.

All patients classified as CAD or with another clinical cause for their presentation can be seen as noise or errors in the querying process. A total of 23 patients with CAD were found in the DBC queries. In two cases incorrect registration was the cause of this. In one case PTCA was not registered and in the other case stenting was not registered. However, when further investigating registration around the procedures, it is possible to correctly identify that these interventions were carried out. The stents used during a procedure are registered in the tab "Used articles" and "PCI" in EPDV. These could be added to the query, to filter out patients where stenting was not correctly registered. For the case where PTCA was not registered this approach would not work, as this is not noted in the previously mentioned tabs. However, this is noted in report created by the cathlab technician during the procedure. This report is based on predefined text where variables such as balloon size and inflation pressure can be manually adjusted. Due to the use of predefined text, these reports are created in a structured and systematic way. These are characteristics favorable for querying, as it is known where what information will be noted. These additional filters could be used to reduce the number of patients with registrations errors in the query results. A total of 11 patients in the CAD group were treated conservatively. The coronary anatomy can be used to distinguish these patients from patients with nCAD. However, as seen in the results of the queries, this tab is not always filled in completely and patients classified as MINOCA are sometimes registered to have affected vessels. Besides the summary in the coronary anatomy tab, CAD patients with conservative treatment can only be distinguished from nCAD patients based on the free-text reports (consultation letter and catheterization report). Seven CAD patients underwent interventions at other centers, this information is only stored in the consultation letter and in some cases in the heart-team meeting. In 3 cases the patient died during or

shortly after the procedure. If the patient dies during the procedure, this is noted in the cathlab technician report, this can possibly be used to exclude these patients. However, two cases the patient died after the procedure, while significant CAD was present. Again, if the coronary anatomy summary is filled in, these patients could be excluded based on this. Furthermore, patient can be excluded based on the free-text reports. A total of 17 patients had another clinical cause for their presentation. This group is very heterogeneous and therefore harder to filter out. The most common diagnosis was SCAD (5 times), followed by pericarditis and arrythmia's (both 2 times). All these patients were registered under one of the three DBC-codes used for the queries, however based on the EHR data these patients were diagnosed with other conditions.

MINOCA and Takotsubo cardiomyopathy

An important consideration regarding the MINOCA and Takotsubo patients is that for most cases the EHR stated that the differential diagnosis consists of MINOCA and Takotsubo, however, in most cases no definitive diagnosis was made. One reason for this is that patients were transferred back to the referring hospital for further investigation of the cause of disease. New information gathered in another hospital is often not available in EPDV. Another cause is that further imaging is required to diagnose Takotsubo, namely echocardiography or left ventriculography. (135) In this analysis the choice was made to classify a patient as MINOCA, when this was the most likely diagnosis, based on the EHR.

Text-based queries

The text-based queries were carried out on the discussion and conclusion tab of the consultation letter and on the catheterization report. After duplicates and patients that were already found with the DBC queries were removed, a total of 35 additional patients were identified. Some patients were returned multiple times by the text-based queries. These duplicates originate from the fact that the queries return every letter or catheterization report where the query terms are found. With consultation letters it is customary to summarize conclusions from previous letters, to give an overview of the patient's history and status. Due to this practice multiple letters per patient can contain this term. This is also the cause of the query returning patients that underwent ICA before 2019. If a patient underwent ICA in 2017 and the diagnosis MINOCA was made, it is possible that this term is repeated in letters after 2019. The 18 patients that underwent ICA before 2019 should not be seen as additional patients, as they fall out of the scope of the queries. This leaves 17 additionally identified patients. One patient did not show up in the DBC queries due to a history of CABG and one patient in the text-based query was found on text referring to cerebral infraction and not a myocardial infraction. These two patients should not be seen as additionally identified patients, as they do not meet the conditions. From the other 15 patients, two did not show up in the DBC queries due to incorrect registration of procedures. These two can be seen as additionally identified patients.

Incorrect use of the term MINOCA is also causing interference. As described in the results section, in one letter a wrong definition of MINOCA is used. Furthermore, patients that have a history of CABG or PCI can, by definition, not have MINOCA. The other 13 patients did not show in the DBC queries, because they were labeled with other DBC-codes. Five patients were labeled with DBC-code thoracic complaints e.c.i. For two of these patients the used DBC-code is incorrect. For one of these patients the referral letter from another hospital states that the patient was suffering a STEMI. The other patients should be labeled as NSTEMI, as the patients satisfies the requirements for this diagnosis and in the letter, it is concluded that the patients was suffering a myocardial infarction. The other three patients with DBC-code thoracic complaints e.c.i. were correctly labeled. Two of these patients can possibly be classified as INOCA. However, not enough information was present to exclude non-cardiac

causes for the presentation. One patient was labeled with the DBC-code pericarditis and one with endocarditis, both with MINOCA in the differential diagnosis. Based on the infection parameters and presentation, pericarditis and endocarditis were more likely. One patient was labeled with the DBC-code supraventricular arrythmia. This patient showed up in the text-based query as this patient suffered a STEMI elsewhere. This was repeated in the letter, however the care given in the LUMC was related to the arrythmia. The other five patients were labeled with the DBC-code stable AP. For one of these patients the letter from the referring hospital stated that this patient was at that moment suffering a STEMI, so the used DBC-code was incorrect. This patient was found 2 times, one time in the query with "MINOCA" and one time with the query "Gepasseerd stolsel". So, 4 unique patients with DBC-code stable AP were found. Another patient labeled with stable AP, was suffering a STEMI, as stated in the letter from the referring hospital. This patient was labeled incorrectly. The remaining two patients had a slight rise in cardiac troponin in other hospitals. Based on their presentation the label stable AP was correct. One of these patients underwent a PCI before, so MINOCA or INOCA were not possible. The other patient possibly suffered from INOCA, however not enough information was present to exclude a non-cardiac cause. In total, 6 additional patients were identified with the text-based queries on the consultation letters (4 incorrect DBC-codes and 2 incorrect procedures). However, when all registration would have been done correctly, all 6 would have shown up in the DBC queries. All three queries resulted in 35 additional patients, after duplicates and patients already found with the DBC queries were removed and only 6 of these patients are in the intended group. Based on this, this way of querying is not very efficient. A lot of noise (incorrect patients) are found and manually checking the EHRs is essential. No hits were found in the catheterization reports. The reason for this might be that this report is only used to describe the things found during ICA and the things done during PCI. Further interpretation of the results is then described in the letter, rather than in the catheterization report.

Overall, it becomes clear that querying EHRs based on DBC-codes for creation of a MINOCA data set is not feasible. It is possible to find the right cases by manually inspection the documentation on every patient. However, this is approach is time consuming and therefore not scalable to the creation of a sufficiently large dataset. Besides this, the yield of queries is not very high. One of the main reasons for this is that classification systems for diseases are created for well-defined and (almost) fully understood clinical entities. This is especially true for DBC-codes, as these codes were created for healthcare insurance and monitoring purposes. Due to this nature, clinical entities that are still very much under investigation, like MINOCA, are not accurately represented by these codes. These types of clinical entities are exactly the subjects that could benefit from ML, to increase understanding and gain new insights on the processes involved. Something else that comes to light is that the DBC-codes are not always a good representation of the actual patient and that human errors in registration are not uncommon.

On the other hand, for clinical entities that are more strictly defined it could be feasible to do queries on local EHR systems in way similar as done here. It is shown here that filtering on procedural registrations can be used to refine query results and that DBC-codes can be used as a coarse searching method. However, further research on this topic is needed.

Another interesting and promising direction for further research on querying EHRs is natural language processing (NLP). In NLP, AI is combined with linguistics. It can be used to analyze structured and unstructured (free-text) data from EHRs and to structure free-text reports into a format that can be processed by ML models. (136) Furthermore, NLP has been successfully used in research setting for accurate phenotyping of complex disease and identifying specific patient populations. (137) In a study

by Redman et al. (2017) NLP is used for the development of an algorithm that can accurately identify the presence of fatty liver disease based on various imaging reports. (138) This might be an interesting direction for querying MINOCA patients, as information from multiple sources is often needed to determine if a patient has MINOCA. Developing a NLP model for this purpose could result in a scalable approach for constructing datasets for ML models.

4.5 Conclusion

Using DBC-codes for querying EHRs to identify and select MINOCA patients is not a feasible approach to create sufficiently large datasets for ML. This is partially caused by the nature of DBC-codes. They are created for health insurance and registration purposes and do not always accurately describe the actual status of the patient. These codes are created late in the process of defining diseases and syndromes. Hence, there are no specific DBC-codes for clinical entities that are still very much under investigation. The DBC-codes can be used for a coarse selection of MINOCA patients, to gain more insight in the distribution and characteristics of MINOCA cases. Furthermore, in this process it is shown that querying with conditions like underwent procedures can be used to refine results. This type of approach might be feasible for querying more strictly defined patient populations. An interesting and promising direction for querying EHRs to select MINOCA patients in NLP. This approach should be further investigated.

5. A CNN for determining lesion significance: a proof-of-concept study

5.1 Introduction

In ICA the coronary arteries are imaged by injecting them with radiopaque contrast, while a continuous X-ray recording is made to assess the lumen of the coronary arteries. At this moment, stenosis severity is typically determined by visual inspection by a cardiologist. This method has several important drawbacks: a significant inter- and intra-rater variability, a high positive prediction bias and the inability to diagnose syndromes with myocardial ischemia without significant stenoses on ICA images. (37, 39) Generally, lesion with a lumen reduction >70% on visual inspection are considered significant. These lesions will mostly be treated with an intervention (PCI or CABG). (22) However, in some cases the visual significance of the stenosis can be unclear. This is especially important in patients with intermediary stenoses (50-90% occluded) and multivessel disease, as there can be a mismatch between anatomical and functional severity of stenoses. (18) In these cases invasive functional assessment can be done during ICA. This can be done with the instantaneous wave-free ratio (iFR) and the fractional flow reserve (FFR). (19) In iFR, a pressure wire is used to measure aortic blood pressure and the blood pressure distal to the stenosis under investigation. When the calculated ratio falls below 0.89, a stenosis is typically classified as significant. (20) The FFR is determined by the ratio of the mean distal coronary artery pressure to the means aortic pressure during maximum hyperemia. A stenosis is classified as significant when this ratio drops below 0.80. (21) An objective method for automated stenosis severity prediction on ICA images could possible improve diagnostic efficiency, while reducing the positive prediction bias. (116) A currently available computer-assisted method for determining stenosis severity is quantitative coronary angiography (QCA). However, this method also has several drawbacks. The process of accurately selecting the right frame for analysis and correctly annotating the region of interest is time consuming relative to the busy workflow of cardiac catheterization. Furthermore, QCA is carried out on one frame, which can cause problems with eccentric stenoses in some situations, similar to visual ICA inspection. (72, 139)

3D-QCA possibly offers a solution for this problem, as multiple ICA images are combined to create a 3D reconstruction of the coronary arteries. However, 3D-QCA is dependent on the quality of ICA acquisition, as the appropriate ICA images and variables might not always be obtained, currently limiting clinical application. (140) Another technique aiming to overcome the problem of eccentric stenoses in QCA, is the quantitative flow ratio (QFR). In QFR the FFR is approximated, based on a 3D reconstruction of the coronary arteries combined with the estimated contrast flow velocity from standard ICA images. (141) In this way FFR values can be approximated without the need for invasive measurements. While results are promising, this technique is not applicable to patients with ostial lesions or patients that underwent CABG before. Furthermore, similar to 3D-QCA, QFR is dependent on a high imaging quality. In a study by Peper et al. (2021), approximately 15% of patients could not be assessed by QFR due to lack of good quality images from the right angles. (142)

Automated software that can be trained beforehand, resulting in fast results when applied in the workflow, would be desirable. Furthermore, relating different aspects regarding the patients to each other, such as acquired images, history and complaints could be valuable for optimizing diagnostics and treatment. A technique that potentially has these characteristics is deep learning. DL is a promising technique for the automated analysis of images. A specific subset of DL algorithms, CNNs are seen as the state-of-the-art algorithms for computer vision tasks and learning spatial features. This technique can simplify training by learning predictive features directly from images. (143) As mentioned in Chapter 2, several algorithms for predicting lesion significance on ICA images have already been

developed. However, the specific data preprocessing steps are not or briefly mentioned in these articles. Furthermore, most classification tasks are performed on single images and not sequences. Possibly the combination of subsequent images, which shows things like contrast movement, could lead to novel insights for diagnostics in ICA regarding yet poorly understood clinical entities such as MINOCA.

In this chapter a DL algorithm was developed for classifying ICA images as significant or insignificant CAD. Based on the perceived potential by the interviewed experts in Chapter 2, developing a ML algorithm for determining lesion significance of intermediary stenoses would be a good step to demonstrate the potential of AI in a way closely related to the clinical practice. However, this is not done with the purpose of developing a generalizable model for stenosis assessment, but to investigate all the steps necessary to go from data stored in EHRs and severs to a DL model. Furthermore, it will be investigated if using multiple subsequent images as input is a feasible direction for research. These things combined can be seen as a step towards developing algorithms for the analysis of more complex clinical entities such as MINOCA.

5.2 Methods

Dataset creation

Data access and querying

The data for the creation of the dataset was retrieved from the EHR system in the department of cardiology in the LUMC (EPD-Vision, as described in Chapter 4). To query the system a SQL-query was created. This query returned all patients that underwent functional assessment during ICA, based on the procedural registration tab in EPDV. The output of this query consisted of patient identification numbers, date of catheterization and if PTCA and stenting were carried out. The query was carried out on all data acquired in 2019-2020.

Data refinement and labeling

For all selected patients the EHRs and images were manually inspected to determine for which vessels the functional assessment was carried out, if this vessel was significantly narrowed according to the functional assessment and if intervention was carried out. Based on this, case labels were determined. If the assessed vessel had a significantly reduced iFR or FFR, the image was labeled as significant CAD. If the assessed vessel had a not (significantly) reduced iFR or FFR the image was labeled as insignificant CAD. The functional assessment was seen as ground truth and vessels without a ground truth were not included. Furthermore, during this inspection the presence of other exclusion criteria was done.

When the left coronary system is imaged, consisting of the LM, LAD and RCx, it is possible that significantly and insignificantly affected vessels are visible at the same time. It is possible that a stenosis is classified as evidently significant on visual inspection. In that case no functional assessment is done, hence no ground truth is present. When this is combined with an insignificantly affected vessel, based on functional assessment, this means that in one image both vessels are visible. These cases are excluded from the dataset, as this will cause errors in model training. Other reasons for exclusion are the presence of implanted devices and materials, such as pacemakers, sternal wires, prosthetic valves and other surgical materials. Patients that underwent CABG before imaging were also excluded, as they have an abnormal anatomy. When there are no image sequences without a wire present, patients are also excluded, as this might act as a confounder. Patients that are evaluated for problems other than CAD, such as aberrant coronaries, are excluded, as different protocols are used in these cases. Movement of the X-ray tube during image acquisition is also a reason for exclusion, as this might cause problems when analyzing image sequences. These cases are excluded to prevent the model from fitting

on confounding factors. For all included patients, the vessel of interest was noted. For each vessel, the imaging sequence with the optimal angle was selected. Table 1.1 in Chapter 1 shows which angle corresponds to which vessel. From each selected sequence, the frame best suited for analysis was selected. This frame corresponds to the frame with the least foreshortening and best contrast filling, similar to the frame used in QCA. (139)

The result of querying, data refinement and labeling was a list of patient identification numbers, date of catheterization, ground truth (significant or insignificant), the number of the selected sequence and the number of the optimal frame.

Data transfer and de-identification

To retrieve the correct image sequences from the CURAD server on which they were saved, the patient identification number, data of catheterization, series number and image number were needed. All images are stored on the local server in a DICOM format. These files were named according to the following format: patient identification number, study date, series number and image number. In the LUMC it is customary to save each image sequence as an independent series, with a unique series number. However, in some cases it is possible to have multiple sequences under one series number. This is the case when images are acquired with a bi-plane set-up, in which two x-ray tubes are used simultaneously to reduce the use of contrast fluid. This is also the case when the operator takes a screenshot of a specific frame in a sequence during the procedure. This screenshot is then stored under the same series number as the whole sequence. To retrieve all DICOM files a SQL query was constructed. All retrieved files were stored on a safe network location in the LUMC, which can only be accessed with appropriate clearance, as the DICOM files contain sensitive patient information. The files never left the internal environment of the LUMC, so no de-identification was applied.

Transformation to machine readable format

The data needed for the training of the model are the imaging data (grayscale values per pixel) and the labels. As all files are stored as DICOM files, containing all sorts of metadata, the pixel data was extracted. A Python script was created to convert the DICOM files to PNG, where each frame resulted in a single PNG file. A schematic overview of the data transfer and transformation is shown in figure 5.1. All data was resized to 512x512 frames, as size varied in the database. Furthermore, all images were transformed to an 8-bit grayscale representation, to ensure a uniform format along all used data. For the training and validation of the network all pixel values were normalized between 0 and 1. This was done to ensure that each input has a similar data distributions, as X-ray devices generate device-specific pixel distributions. The data used in this study was acquired with three different devices. 152 cases were acquired with a Philips Azurion system, 32 with a Philips Allura Xper system and 59 with a Siemens AXIOM-Artis system.

Fig. 5.1 Schematic representation of data transfer and transformation

Exploratory data analysis

The distribution of the two labels and the distribution of the vessels included were determined to create an overview of the dataset

Organize data

After all data was transformed and saved in a structured way, the data was split randomly, while maintaining case balance. This was done by creation a folder with the frames for cases labeled as significant CAD and a folder for cases labeled as insignificant CAD. From each folder the cases were randomly divided to be a part of the training or the validation set, with a ratio of 80:20.

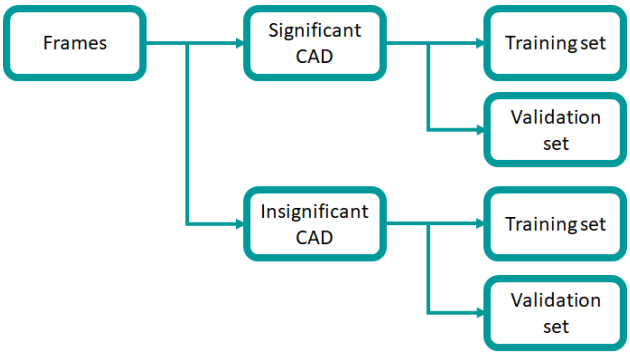


Fig. 5.2 A schematic representation of the organization of the data for training and validation.

Networks

Two networks based on the VGG16 architecture were created and evaluated. VGG16 is a CNN created by Simonyan and Zisserman. The VGG16 model reached 92.7% top-5 accuracy in ImageNET, a large dataset of more than 14 million images belonging to 1000 classes. (144) Batch normalization and dropout layers were added to the networks. Dropout layers reduce chances of overfitting. The dropout layer randomly inactivates neural units during the training, at a specified rate. Batch normalization is known to improve the accuracy of CNNs without any side effects. In batch normalization the layers inputs are normalized by setting the mean of the inputs to 0 and the standard deviation to 1. (145) According to a study by Garbin et al. (2020) adding dropout layers to CNNs should be carefully considered and differences in performance before and after adding these layers should be monitored closely. (145) For this reason only one dropout layer, with a dropout rate of 0.2, was eventually added to both networks. For both networks, the schematic representations are shown in figure 5.3a-b.

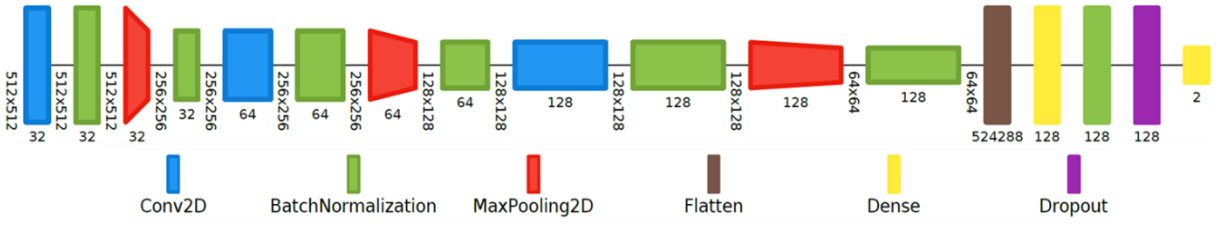


Fig. 5.3a Schematic representation of network 1. The input dimensions for each layer are shown in vertical text and the number of filters the layer will learn from in horizontal text.

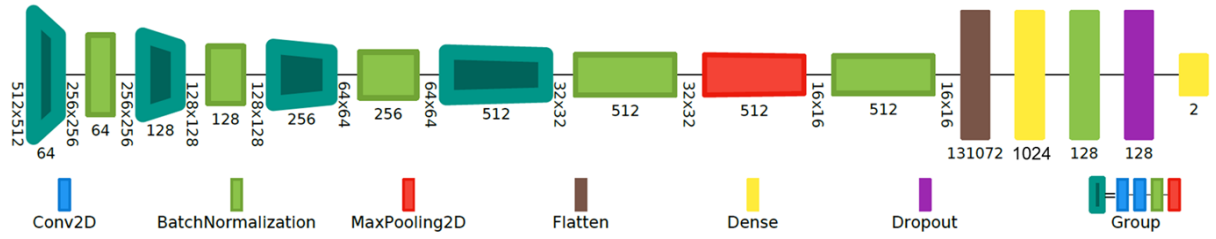


Fig. 5.3b A schematic representation of the network 2. The input dimensions for each layer are shown in vertical text and the number of filters the layer will learn from in horizontal text. To create a clearer visualization the combination of two convolutional layers, a batch normalization and one max pooling layer is shown as one component.

Network 1
The network, as shown in figure 5.3a, consists of 67,204,226 parameters, of which 67,203,074 are trainable and 1,152 are not trainable. This network was also evaluated without batch normalization and a dropout layer, this results in a total of 67,201,922 parameters, which all are trainable.

Network 2
The network, as shown in figure 5.3b, consists of 138,918,850 parameters, of which 138,911,938 are trainable and 6,912 are not. Without the batch normalization and dropout layer the network consists of 138,905,026 parameters, which all are trainable.

Algorithm training and evaluation
Data augmentation is an important preprocessing methods that is known to be effective for training highly discriminative DL models. (146) CNNs generally contain millions of parameters and thus require high amounts of data for reaching a high performance. In data augmentation the amount of data is artificially increased by applying random transformations to images. To increase the amount of data used for training data augmentation was applied in this study. The transformations applied were shifts in pixel intensities, rotation of the images, shifts in height and width and shearing. The applied values for each transformation can be found in Appendix 3. Examples of augmented images are shown in figure 5.4.

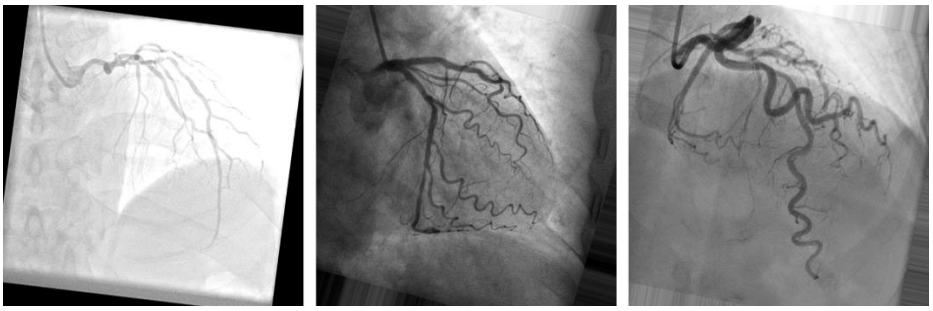


Fig. 5.4 Examples of randomly augmented images

Training was carried out on 512x512 images for both networks. A batch size of 5 was used in network 1 and a batch size of 1 in network 2, due to memory constraints. The Adam optimizer was used with a learning rate of 10E-5. Binary cross-entropy was used as loss function and accuracy as performance metric. An early stopping module was implemented, to reduce unnecessary training time. When the validation accuracy did not improve during 40 epochs, the training was stopped. The maximum number of epochs was set to 100 and all data was analyzed during every epoch. The training loss and accuracy and the validation loss and accuracy were plotted to summarize the training process and performance. For each network, a confusion matrix was calculated on the validation set, from this classification accuracy, recall (also known as sensitivity), specificity, precision and the F1 score are calculated. The equations for these metrics are shown in Appendix 3.4 Insignificant CAD is seen as the negative class and significant CAD as the positive class.

Table 5.1 Confusion matrix

	Predicted positive	**Predicted negative**
Actual positive	True positives	False negatives
Actual negative	False positives	True negatives

Furthermore, Gradient-weighted Class Activation Mapping (GRAD CAM) visualization were created to give more insight into the characteristics of the network. A GRAD CAM visualization gives a coarse visualization map that highlights regions important for the prediction made, based on the final layer of a network. (147) A stratified k-fold cross-validation was carried out to estimate the skill of the network to unseen data. Results can indicate overfitting and a lack of generalizability. Generally a k between 5-10 is used. (126) Here the choice was made for k = 5.

Analysis of multiple subsequent frames

To train the networks on multiple subsequent frames, three subsequent frames were merged in one file for all cases. These files were saves as RGB images, where each original frame is represented by one channel. The frames used for this were the selected frame used in the training of the other models, one frame before the selected frame and one frame after the selected frame. The structure of the networks had to be slightly altered to accept three channel input images instead of one channel (grayscale) images. An example of the created three channel image is shown in figure 5.5.

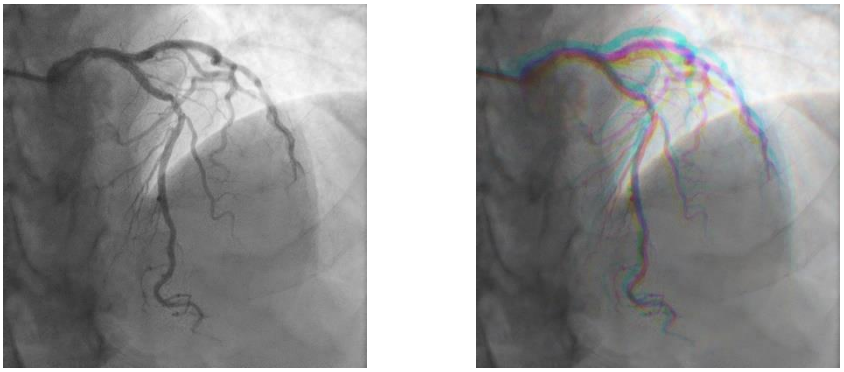


Fig. 5.5 Left the original frame used for training the networks, right the three-channel file.

Implementation details

All scripts and models were implemented in Python (3.7.10) using the TensorFlow library (2.0.0). Conversion of the DICOM data to PNG files was carried out with Python 3.6.13, as the decoding of the compressed DICOM files needed a package that was not available in Python 3.7.10. All python scripts are shown in Appendix 3.3. The code for the manual stratified k-fold cross-validation and the three channel images are not shown here, as these are just slightly different than the code shown. These scripts are saved on the workstation. Training and validation were carried out on a workstation with Intel Xeon E5-1620 V4 3.5 GHz CPU, 128 GB RAM, and a NVIDIA GeForce GTX 1080 GPU.

5.3 Results

Dataset creation

The query resulted in 450 patients, 185 patients underwent stenting and PTCA, 4 only PTCA, 15 only stenting and 239 did not underwent any intervention. In 274 (61%) cases the intervention status based on the procedural registration was equal to the interventional status of the stenosis that was assessed with functional assessment. A total of 207 cases were excluded: 86 patients had another significant lesion visible in the image, for which no iFR assessment was done. A pacemaker was present in 27 cases, 17 cases were evaluated with the a protocol not aimed at detecting CAD (aberrant coronary anatomy), for 13 cases there were no appropriate images without a wire present, in 7 cases the reports were inconclusive, in 3 cases sternal wires were present, in 1 case a prosthetic valve was present, in 1 cases surgical clips were present, 10 cases underwent CABG before image acquisition, in 30 cases the x-ray tube was moved during acquisition, in 9 cases no iFR or FFR measurement was carried out, in 1 case no ICA was carried out, in 1 case the signal-to-noise ratio was very low and in 1 case SCAD was diagnosed.

A total of 243 cases were included of which 87 were classified as significant CAD and the other 156 as insignificant. In 46 cases the vessel of interest was the RCA, in 154 cases the LAD and in 43 cases the LAD. A total overview of the case selection is shown in figure 5.6

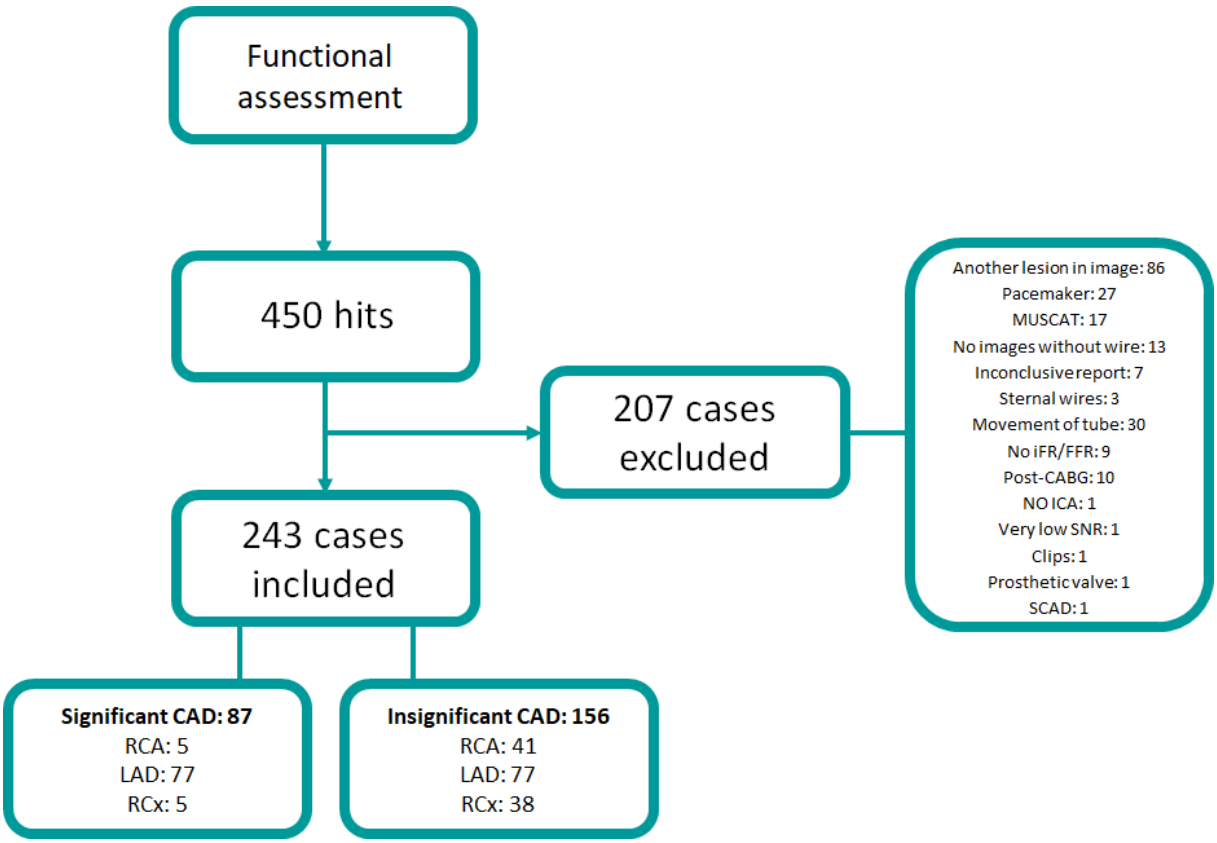


Fig. 5.6 Case selection for the creation of the dataset.

Data transfer

All included cases were collected from the CURAD server. The query for 243 patients, resulted in 252 DICOM files. In 9 cases more than one image sequences was acquired per series number. In three cases bi-plane acquisition was the cause of this, in the other 6 cases it were screenshots.

Minority class oversampling

During early training of the networks, it became clear that the case imbalance was significantly influencing classification performance. The networks reached an accuracy of 0.64 on the validation set when trained with the imbalanced dataset. Approximately 64% of all cases was labeled as insignificant during dataset creation. When further analyzing the confusion matrix for these networks, it became clear that all cases were classified as insignificant, hence the accuracy of 0.64. To prevent this random minority class oversampling was applied, resulting in a 50/50 ratio for the classes.

Network Performance

A total of six network configurations were trained: network 1 without batch normalization and dropout (N1), network 1 with batch normalization and without dropout (N1_BN) and network 1 with batch normalization and dropout (N1_BN_DO). The same three configurations were tested for network 2 (N2, N2_BN, N2_BN_DO) During network training the loss and accuracy on the training and validation set were monitored as learning curves. The learning curves for two networks are shown in figure 5.7a-b. In these curves the loss and accuracy for the training and validation set were calculated each epoch and the network weights of the configuration with the highest performance on the validation set were saved.

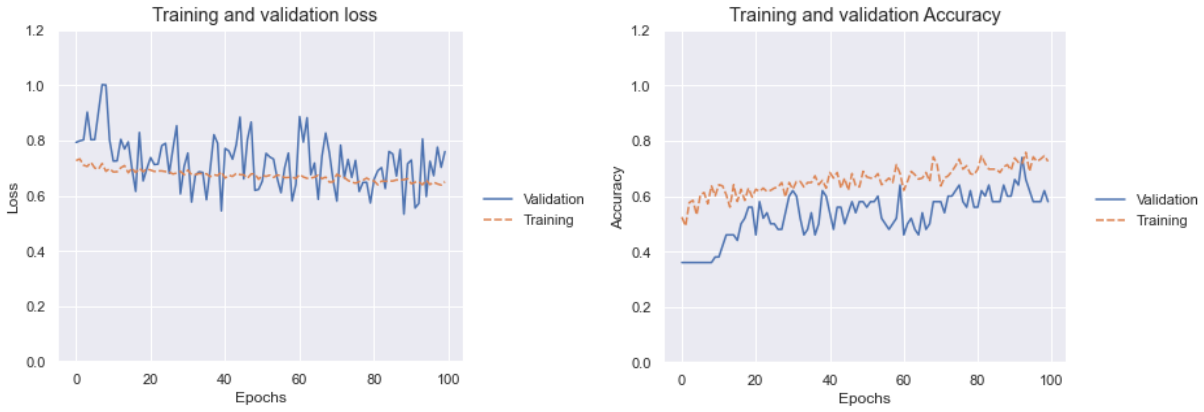


Fig. 5.7a Learning curves for N1_BN_DO with on the left the loss during training and validation and on right the accuracy.

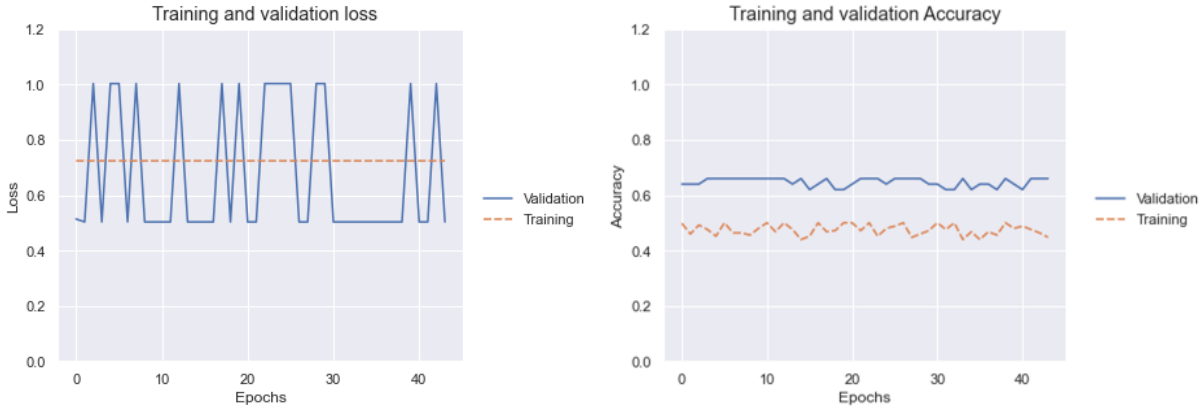


Fig 5.7b Learning curves for N2_BN with on the left the loss during training and validation and on the right the accuracy.

The confusion matrices are shown in Appendix 3.4. The computed metrics for the six network configurations are shown in table 5.2 and 5.3. The highest precision is achieved by N1 and N2 with a value of 0.714. The highest recall (0.889) was reached by N1_BN_DO, the highest specificity (0.938) by N2, the highest F1-score (0.711) by N1_BN_DO and the highest accuracy (0.760) by N1.

Table 5.2 Metrics calculated on validation data for network architecture 1

	N1	N1_BN	N1_BN_DO
Precision	0.714	0.600	0.593
Recall	0.556	0.667	0.889
Specificity	0.737	0.667	0.568
F1-score	0.625	0.632	0.711
Accuracy	0.760	0.720	0.740

Table 5.3 Metrics calculated on validation data for network architecture 2

	N2	N2_BN	N2_BN_DO
Precision	0.714	0.522	0.636
Recall	0.278	0.667	0.389
Specificity	0.938	0.656	0.875
F1-score	0.400	0.585	0.483
Accuracy	0.700	0.660	0.700

Grad CAM

For all models Grad CAM visualizations were made, to produce a coarse insight on what regions of the input images are important for classification. Several representative visualizations are shown in figures 5.8-9. The colormap used for visualizing the heatmap was a jet colormap, meaning that the importance of an area increases from blue to red. The first three images of figure 5.8 are based on N2_BN_DO and the second three on N2. In figure 5.9 the first three images are based on N2_BN_DO trained with a learning rate of 10^{-5} (like all other networks) and the second three are also based on N2_BN_DO but trained with a learning rate of 10^{-3}.

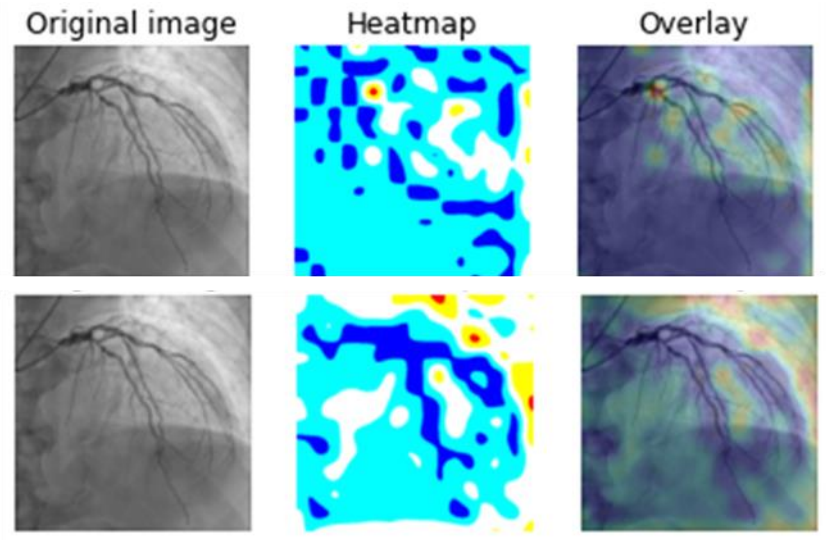


Fig. 5.8 Grad CAM visualization of N2_BN_DO (top) and N2 (bottom).

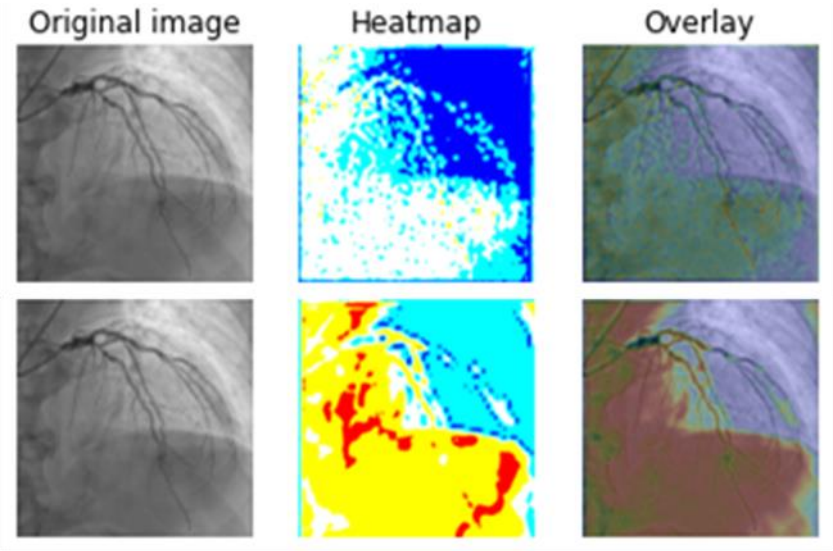


Fig. 5.9 GRAD CAM heatmaps combined with the original image. The top heatmap is based on N1_BN_DO (learning rate 10^{-5}) and the bottom one also on N1_BN_DO but with an increased learning rate (10^{-3})

K-fold cross-validation

A 5-fold stratified cross-validation was carried out for N1, N2, N1_BN_DO and N2_BN_DO.
The average accuracy for N1 was 0,662 with a SD of 0,0424. The average accuracy for N1_BN_DO was 0.672 with a SD of 0.026. For N2 the average accuracy was 0.692 with a SD of 0.033. For N2_BN_DO the average accuracy was 0.698 and the SD 0.030.

Three-channel images

The three-channel image dataset was used to train the networks N1, N1_BN and N1_BN_DO. Due to memory constraints is was not possible to train the N2 networks with this data. The computed metrics are shown in table 5.4. N1 reached the highest precision (0.667) and F1-score (0,651), N1_BN_DO reached the highest recall (0.667) and specificity (0.657). N1 and N1_BN_DO both reached an accuracy of 0.700. The Grad CAM visualization for N1_BN_DO is shown in figure 5.10.

Table 5.4 Metrics computed on the confusion matrices for the three models trained with three-channel images.

	N1	**N1_BN**	**N1_BN_DO**
Precision	0.667	0.500	0.571
Recall	0.636	0.579	0.667
Specificity	0.600	0.64	0.657
F1-score	0.651	0.537	0.615
Accuracy	0.70	0.620	0.700

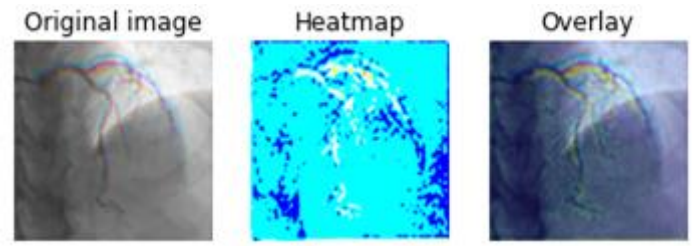


Fig. 5.10 Grad CAM visualization for the final layer of the N1_BN_DO network with a three-channel images as input.

5.4 Discussion

In this study a strategy for creating a dataset for the creation of a DL model from an EHR system was developed. This includes querying an EHR system, case selection, refining the gathered data, labeling the data, transferring the data and all the necessary preprocessing steps required for the training of a DL model. Furthermore, two DL networks were developed to assess the feasibility of this approach. These models were developed for the analysis of single ICA frames, however, a small step towards analyzing multiple subsequent frames was made. The gathered information can be used as starting points for further research on trying to improve diagnostics and therapy for clinical entities such as MINOCA in interventional cardiology.

Dataset creation

The dataset creation started with a query for patient that underwent functional assessment in 2019-2020. The choice for this requirement was made to have an objective ground truth on lesion significance. In this way the influence of subjective choices by operators in visual assessment is minimized, as functional assessment provides an objective quantification of the stenosis severity. However, this strategy also creates a bias towards the selection of cases with images that are difficult to interpret, as an operator generally chooses for functional assessment when visual inspection is inconclusive or unclear. (18) This means that most cases included in the dataset are these difficult cases, possibly making it more challenging for the network to learn discriminative features. Possibly, including straightforward cases in the training process can improve performance by making the network focus more on the typically narrowed parts of the coronary arteries. The raw querying results were not directly useable, as eventually 46% of all found cases was excluded. Furthermore, the interventional status based on the querying results was an accurate depiction of the lesion significance of the functionally assessed vessel in only 61% of the cases. After assessing all cases for the eligibility criteria and excluding non-eligible cases, the sequence with the right acquisition angle for the vessels of interest needed to be selected. After the right sequence was found, the right frame, with the least foreshortening and best contrast filling had to be selected. Due to these criteria, manual inspection of EHR data, such as consultation letters and procedural reports, and manual inspection of the acquired images was needed. This was a time-consuming process, while the dataset size is still small compared to similar studies, as shown in table 2.6.

Network structure and performance

The networks created in this study were based on the well-known VGG16 network. (144) However, no full VGG16 network was implemented. On reason for this was the memory constraint of the used GPU. As a full VGG16 network is large and the images used for training are also quite large (512x512). Furthermore, in other studies on classification of ICA images it was found that CNNs with fewer layers generally perform better than larger networks. This suggests that that stenosis characterization relies more on low-level visual features, which are typically recognized by the first several layers of a network. (72, 116) This is in line with the way stenoses are characterized in visual inspection, where severity is determined by estimating the narrowing of the arteries, which can be seen as curve-like image features. (72) This might be the reason that the networks with only three convolutional layers have a slightly better performance than the larger networks (containing 8 convolutional layers). Training and validation was done on small datasets and the ICA images contain several background structures, which makes the networks prone to fitting on confounders. (114) To reduce this, batch normalization and dropout layers were added to both networks. In the 5-fold cross-validation the performance of the networks with batch normalization and dropout layers is slightly higher and the standard deviation a bit lower than that of the networks without these layers, however, these differences are small and could also come from the stochastic nature of the process. The Grad CAM

visualizations before and after adding batch normalization and a dropout layer, shown in figure 5.8, show that before these layers were added the coronary arteries show up as blue (not important) on the heatmap, while important spots (in red) lie outside the vasculature. After applying batch normalization and dropout regularization the importance of the area that previously showed up red is decreased and an important spot has showed up around the distal part of the left main branch and the ostia of the LAD and RCx. In this patient a hemodynamically significant stenosis (iFR = 0.82) was identified in the proximal LAD and a stent was places here. This might indicate that the network was able to detect a stenosis here. However, it cannot be excluded that this is based on coincidence.

The two sets of learning curves shown in figure 5.7a-b show the two types of learning curves encountered during the training of the networks. The curves in figure 5.7a show a slight decrease of loss and a slight increase of accuracy during training on both the training and the validation set. However, all curves have an unstable course. This indicates that the dataset does not provide sufficient information to learn how to solve the problem and that the validation set does not provide sufficient information to evaluate the models generalizability. (126) The small amount of data used is likely the cause of this behavior. The set of learning curves shown in Figure 5.7b show a different course: the training loss does not change during training, this indicates underfitting. Underfitting means that the model was unable to learn from the training dataset. This can be caused by the not having the suitable capacity for the complexity of the dataset. (126) However, these curves only occurred in the networks based on 8 convolution layers and not in the models with 3. Another cause can be the presence of too much noise. ICA data is known to contain high levels of noise. This also becomes clear in the Grad CAM visualizations, as often background pixels are given a high level of importance, likely due to the presence of noise. This can be noise resulting from image acquisition, however the presence of other structures such as the diaphragm can also be seen as noise, as these structures do not contain any information on the coronary artery status of a patient.

It is not straightforward to conclude which model performed the best, based on the calculated metrics, as every metrics focuses on a specific ratio between classification results and there is often a tradeoff between metrics. This is shown by the metrics for network N2, it has a high specificity, meaning that a large proportion of the negative cases was identified correctly. However, the recall value is the lowest of all networks. This means that this network tended to classify cases as negative. Accuracy is generally the most intuitive metric, however, this metric can be deceiving. Network N1 showed the highest accuracy, however the recall is relatively low. This means that overall, this network classified the most cases in a correct way, as compared to the other network. The recall value of 0.556 indicated that only 56% of all cases with a positive label were classified as such. The F1-score shows the harmonic mean between recall and precision. Network N1_BN_DO has the highest F1-score and the highest recall, the precision however is the second lowest. When selecting the best performing network, it can be insightful to consider the implications of the different metrics, especially in a medical context. In the context of ICA and deciding on whether an intervention should be carried out a false negative means that a patient would not receive treatment, while there might be a risk for adverse events such as a myocardial infarction. On the other hand, a false positive means that a patient receives treatment while not needing it. This is also problematic, as intervention itself comes with risks and costs a significant amount of money. However, one might have a slight favor towards performing an intervention when not needed than to not performing an intervention when needed, although this is very dependent on the presentation of the patient. When applying this, N1_BN_DO can be seen as the best performing model, as it had the highest recall, highest F1-score and the second highest accuracy.

Limitations

Due to time constraints the networks were not fully optimized and further analysis is still needed. These time constraints were also the reason for the small dataset size. Due to this no test set, a dataset not seen by the network in either training or tuning of the hyperparameters, was created and used for an unbiased evaluation of the network. As the network configuration is not only based on the training, but also the validation set, testing on the validation set can give a result biased towards a higher performance. Another limitation of this study was the memory capacity of the used hardware. Training the large network on the three channel images with a 512x512 image size resulted in an out of memory error. This was also the reason no full VGG16 network was analyzed. This is especially of importance for further research on training network on multiple subsequent frames. The goal of this study was not to create a generalizable model for predict stenosis significance. The results should also not be interpreted in that way. This limits the scope of the study, as it serves as an illustration of how these subjects can be investigated, which aspects should be considered when setting up similar studies and what subjects should be further investigated.

Further research

For further research increasing dataset size would be an interesting direction. With a larger dataset more research can be done on whether the model is actually fitting on features that are regarded as important, such as vessel narrowing, rather than confounders in the images, such as visible background structures. However, as mentioned before, creating datasets with the strategy applied here is a time-consuming process and subjective choices are made in frame selection and interpretation of the reports in the EHR system. Automation of this process could reduce the time needed and decrease subjective choices. For the analysis of EHRs NLP would be an interesting direction and this is not limited to the subject researched here. As mentioned in Chapter 4, NLP could also play a role in the analysis of complex clinical entities such as INOCA and MINOCA. This can be done by analyzing structured and unstructured data from EHRs, but it can also be used to structure free-text reports to a machine readable format, which can be used in ML models. (136) Another part of the process that could be automated is the selection of the sequence with the right acquisition angle and the frame selection. In a study by Cong et al. (2019) a coronary artery – angle view network was developed (based on Inception V3). It was specified which angles should be used for which artery. This network learned to classify the image as RCA of LCA. Next DICOM tags were used to identify the appropriate vertical and horizontal angulations. In this study another network (CNN + LSTM-RNN) was created for frame selection. (114) After this process the selected images were still manually inspected, however, this approach could reduce the time needed for data selection. Similar approaches to automating data selection are used in several other studies identified in the systematic review (Chapter 2). Generating synthetic data from a dataset is another approach for increasing dataset size. A promising approach for this is the use of generative adversarial networks (GAN). GANs are neural networks that can generate synthetic images. In this way data shortage could be reduced. This technique has an advantages over traditional data augmentation techniques, as the distributions of traditionally augmented images have similar distributions as the original ones. (148)

To gain further insight in the classification performance, it would be interesting to analyze the class probabilities generated by the SoftMax layer. In this study the class with the highest probability was assigned as classification results, however, it would be interesting to analyze with what amount of certainty the network assigns a class. Including cases with obvious stenoses or with obviously normal coronary arteries in the dataset for training can also be an interesting direction. This might improve network performance, as the differences between the two classes are more evident and possibly this can help the network focus on the presence of vessel narrowing.

From the GRAD CAM visualizations, it becomes clear that background structures are often part of the classification process. This is caused by the heterogeneity of the ICA images and variations across patients and acquisition parameters. In several studies on stenosis classification on ICA images, a segmentation or detection algorithm is used before classifying the stenosis. (72, 84, 89, 113, 116, 121) In this way the background information is not used while classifying the stenosis and therefore not acting as a confounding factor. Furthermore, in several studies the segments of the coronary arteries are named by a model. (114, 115, 117, 119, 120) This results in a classification result per artery, circumventing the problem of multiple vessels in one image mentioned in the methods section. Currently, most DL models developed for disease classification on medical images consider only pixel values and not clinical context, such as anamnesis, age and medical history. While in practice physicians combine image data with clinical context to come to a conclusion. In other fields were image data is combined with other data sources, it is known that a so called multimodal DL model often achieves higher accuracy than single-modal DL models. (149) Combining ICA images with data from EHRs or imaging data from other modalities, such as CT, IVUS or OCT, would thus be an interesting direction for further development of a clinically applicable DL model. Besides aiding classification, data from other modalities could also be used to automate data selection. Related to the data preparation strategy in this study, for the analysis of ICA with QCA it is generally recommended to select an end-diastolic frame. (139) The data acquired during ICA is stored in DICOM files, these files contain a wide range of data regarding the patient and image acquisition. Generally, the DICOM files containing ICA images also contain ECG recordings acquired during image acquisition. Possibly, this information can be used to automate frame selection.

In this study an early attempt was made to use multiple subsequent ICA frames for the classification of CAD. This was done by merging three grayscale frames to one three channel image and using this for training and validation on the same networks as the single grayscale images, based on 2D convolutions. However, for further research it would be interesting to tune a network specifically for this purpose. A smaller variant of the VGG16 network, such as the VGG11 network, can possibly be used as a starting point. Smaller networks might be of interest, as stenosis classification on ICA images possibly depends largely on low-level image features. Furthermore, training smaller networks requires less GPU memory. It might also be interesting to investigate the application of DL models other than CNNs on a stack of ICA frames, such as LSTM-RNN for the analysis of time-based dependencies. In a study by Ma et al. (2019) optical flow estimation is used for segmentation of ICA sequences. Perhaps this technique could be used to further analyze flow patterns in ICA sequences.

5.5 Conclusion

In this study a dataset was created and used for the development of CNNs that classifies coronary artery stenoses as significant or insignificant. The main purpose of this study was not to create a generalizable model, but to investigate what is needed to go from patient data in an EHR system to a DL network. The identified steps were: querying the EHR system for a starting dataset, refining this dataset by applying eligibility criteria, transferring the selected imaging data for the included patients, converting the data to a format suitable for ML and applying preprocessing steps to create a uniform data format. The results of analyzing the dataset show that more data is needed, while the used strategy for dataset creation is not scalable. For this reason, additional research is needed on automating this process. The networks developed in this study show reasonable performance is some cases, however further analysis of this is needed. Furthermore, several subjects for further research on this specific topic and on analyzing complex clinical entities such as MINOCA were identified.

6. A roadmap for the creation of datasets for developing medical ML algorithms in interventional cardiology

6.1 Introduction

As mentioned in Chapter 1, a ML-based algorithm to guide ICA diagnostics and treatment decisions can be a solution to several problems encountered in interventional cardiology. However, despite promising results in both radiology and cardiology, implementation of ML is still limited. (62, 64) Two major factors contributing to this are a lack of well-defined questions and a lack of high-quality annotated datasets. (64, 66, 150)

Current AI research is often not focused on clinical practice and patient outcomes, while this is essential to deliver a positive impact on healthcare. Clinicians should take the lead in defining clinical problems that can benefit from ML models, as they have the best knowledge of problems in clinical practice and will often be the end users of the application. (122) Besides this, the metrics used for evaluating ML models are often not translatable to clinical effectiveness and models are difficult to compare due to large differences in methodology. (66) This makes it hard to identify where ML can improve healthcare and what the actual value of these algorithms is. (65) Another factor contributing to this, is that the clinical workflow and information processing pipeline is not always considered thoroughly, leading to models that do not fit into practice. (122, 127)

Besides the translational problems for the ML models created to clinical practice, the lack of high-quality annotated data for model training is a major limiting factor for clinical application of ML models. (64, 66) A limited amount or low-quality data for training can lead to problems with generalizability, bias and fitting on confounders. (66) This can make a ML model inaccurate and unusable in clinical practice. Lack of knowledge on ML by physicians is also a factor that limits implementation in clinical practice. (66) To overcome these limitations and increase the positive impact of ML on healthcare a systemic approach for defining clinical ML applications and data curation should be applied. In this chapter results from subsequent chapters were combined with scientific literature to create an overview of the steps and considerations needed for this. Furthermore, the translation of these steps to the practice of interventional cardiology was made.

Data hierarchy

High quality datasets are the cornerstone of good performing ML models. The smallest piece of data in a dataset, is a data element or a data sample, an individual piece of data. This can be every form of medical data, from a patients age to an imaging sequence. A data record is a set of related data elements, in the context of the clinical practice of interventional cardiology this is a patient. Multiple data records together from a dataset, typically the data records in one dataset have approximately the same data elements. A dataset can also be part of a larger structure, namely a repository. A data repository, also known as a database, is an infrastructure that hosts one or more datasets for retrieval and querying. (151) A schematic overview of the data hierarchy is shown in figure 6.1.

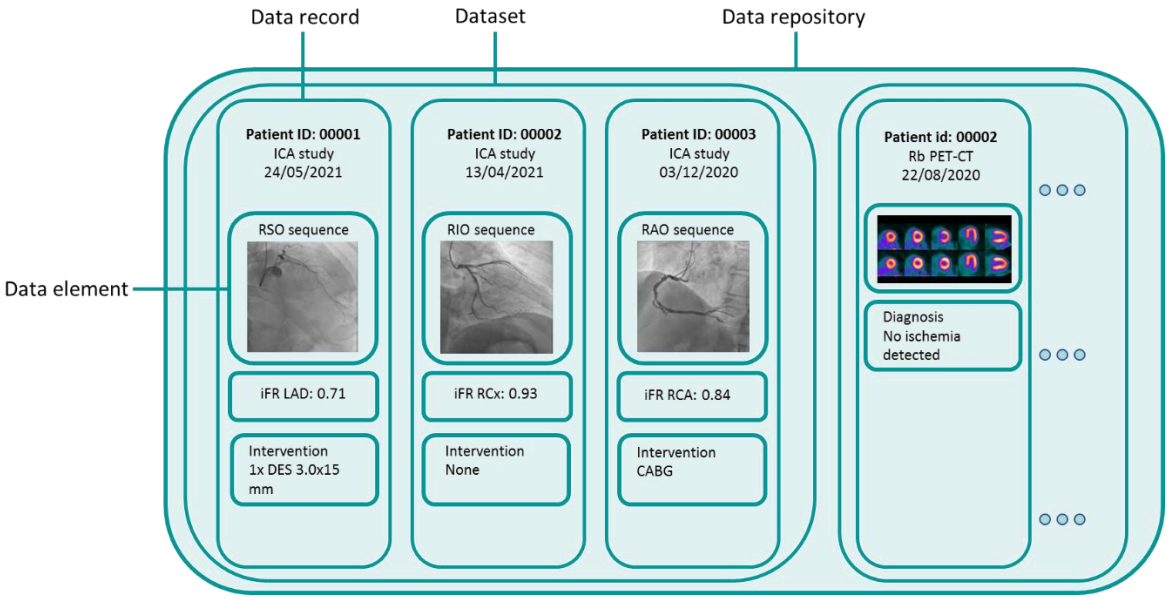


Fig. 6.1 Data hierarchy, a data element is a single piece of data, related data elements form a data record. A collection of similar data records is a dataset and a data repository curates and hosts one or more datasets.

6.2 Roadmap

6.2.1 Framework

To ensure a focus on clinical applicability, the creation of an ML model should start by creating a framework. This should start with defining the clinical problem for which the model will be created. A possible approach for this is interviewing experts, as in Chapter 3. In interventional cardiology, the interventional cardiologists generally have the best knowledge of clinical problems, as they acquire the images, interpret them and carry out interventions when necessary.

Next, the workflow in clinical practice related to the clinical problem should be described. This description should include key moments for decision making and an overview of what data is acquired at what point in the process. This can be used to further specify how the ML model should fit into the workflow. (150) A general overview of the workflow in interventional cardiology, regarding CAD, is shown in figure 6.2. From this overview, a clear difference between the workflow of interventional cardiology and radiology becomes clear. Generally, in radiology image acquisition, interpretation and acting on the acquired information happens in separated steps, often at different times. In interventional cardiology these steps are closely connected to each other, as shown in the workflow overview. Key moments in the process are the choice to perform an ICA, interpretation of ICA images, the choice for additional imaging or testing and the choice of performing or scheduling an intervention. The choice to perform an ICA is based on other diagnostic testing, anamnesis and physical examination, as described in Chapter 1. The choice for additional imaging or testing is made based on information gathered from acquired images and eventually the combination of all this information is used to determine the right strategy for the patient.

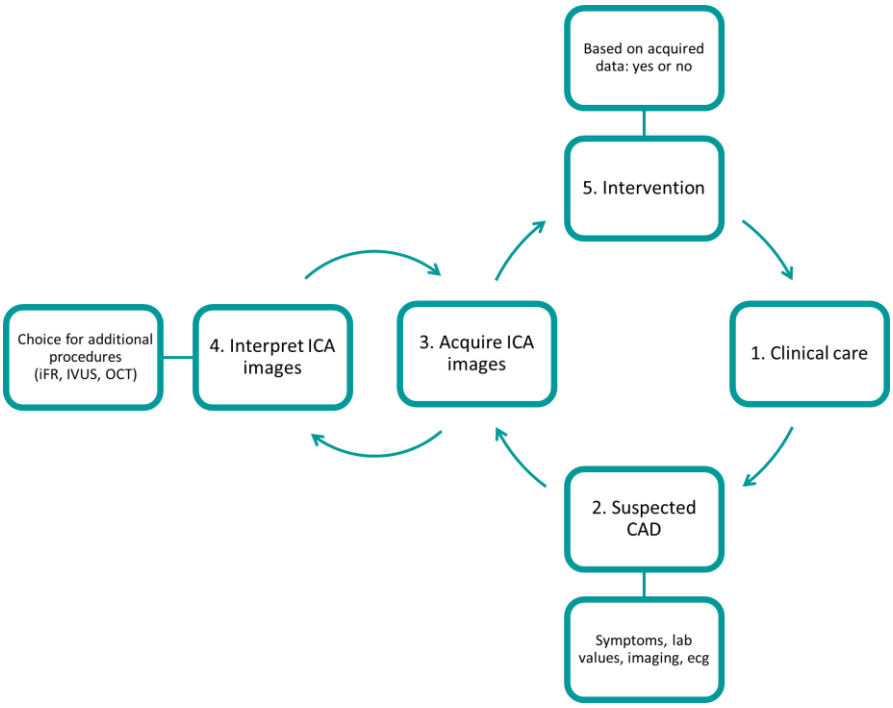


Fig. 6.2 A general overview of the workflow around patients suspected of coronary artery disease.

After identifying a clinical problem that can possibly benefit from a ML model and describing the workflow related to the problem, an early description of the model should be created. A good way of doing this is by creating a use case. (150) A use case is a conceptual model that shows what a system does and with what aim. (152) In a use case all actors (users) interacting with a system are described showing the reasons, causes and way of interacting with the system. (153) A good use case defines a project with a specific aim, a measure of success, a clear description of the users and a rationale for its value. (154) It is also important to describe what data is required for the creation of a model that can solve the clinical problem and which type of models are likely to lead to good results. (155) The aim of the ML model in the use case, is to solve the clinical problem for which the framework is created. To create a concrete and concise overview of the use case for the ML model, the template shown in figure 6.3 can be used. In this template the input data, expected algorithm type, granularity of the model output and the type of output are described. (156)

In interventional cardiology, the input data for the model will likely be images acquired during ICA, possibly combined with other sources of data such as intravascular imaging, iFR data and data from EHRs. The type of algorithms that are expected to perform well are strongly dependent on the input data and the desired outputs. However, as interventional cardiology revolves around imaging data, deep learning models are likely to be the best suited. (30) Specifically, convolutional neural networks (CNNs) are seens as the state-of-the-art algorithms for computer vision tasks and learning spatial features. (143) Recurrent neural networks (RRNs) and specifically long short-term memory (LSTM) networks, are good in capturing dependencies in sequential data and so are well suited for learning time-based dependencies. (157) The granularity of the prediction by the model is likely per-input, as opposed to intra- or inter-input, as in most cases a model would be used to make a prediction for each input image or imaging sequence. However, the other types of granularity can also occur. The output decision is also task dependant. An often used type in computer vision tasks is class, e.g. for predicting stenosis significance the output decision would be class 1 (significant) or class 2 (insignificant). Another commonly used approach is a probability as output, e.g. what is the probability of in-stent restenosis for a specific patient based on the input image.

The definition of success differs between ML models, however, when developing a ML model for clinical practice, the aim will almost always be to improve outcomes for patients. As mentioned by Kelly et al. (2019), performance metrics for ML models should aim to be relatable to clinical application and be understandable for intended users. (66) Currently, this is often not the case. (158) Choosing the right metric is important for both assessing model performance and guidance of modeling. (159) Which metric is best suited depends on the eventual goal of the model and the class distribution in the dataset. Generally metrics can be divided into three categories: threshold metrics, ranking metrics and probability metrics. (160) Threshold metrics quantify classification prediction errors. This group includes metrics like sensitivity, specificity and accuracy. These metrics are often used in the development of classification models for medical applications, as they are straightforward and clinicians are often familiar with the definitions. (66, 161, 162) Ranking metrics focus on how well models perform at separating classes. The classifier needs to predict the probability that a case belongs to a class for this type of metrics. (161, 162) A common ranking metric is the ROC-curve and related to that the area under the curve (AUC). The ROC-curve is a plot of the model behavior by calculating true and false positive rates under different thresholds. The AUC summarizes the ROC-curve in a single value. An AUC of 0.5 indicates the performance of a classifier that randomly assigns classes, also known as a no-skill classifier. A perfect classifier has an AUC of 1.0. (161) Probabilistic metrics are used to quantify the uncertainty of a model's prediction. This is useful when the focus is on assessing not only if a prediction was correct, but also if the wrong class was selected with high or low probability. Common probabilistic metrics are logarithmic loss and the Brier score. (161) These type of metrics can be used to calculate the probability that a patient belongs to a certain class, e.g. what is the probability that a patient will have in-stent restenosis after PCI.

Fig. 6.3 This template can be used to create a very concise visualization of the ML use case. The input data described which data is used as input, the algorithm described what type of algorithm is likely to perform well, granularity describes how granular the prediction of the model should be and the output decision describes what type of output is desired. (156)

When choosing evaluation metrics, it should also be considered what the impact of errors on patients would be. This is best illustrated with an example. When an algorithm is used to diagnose the presence of a pulmonary embolism, a false negative classification will have a much larger impact than a false positive error. So, in this case a high sensitivity would be more important than a high specificity, ensuring that most patients that have the condition are classified as such.

It is also important to consider class balance when selecting a metric. Several metrics work very well in balanced datasets, but lead to serious problems when classes are imbalanced. (163) Accuracy is a commonly used metric, defined as all correctly classified cases divided by the total amount of cases. When this metric is used in a very unbalanced dataset, this will lead to the model classifying all cases as the majority class, as this will result in the highest value for this metric, as shown in Chapter 5. (164) Selection of the right evaluation metric is one of the most important problems in ML. (161) The best way to approach this problem might be to discuss with stakeholders what is important in the context of the model and the environment it will be used in. Furthermore, using several different metrics can give a more complete view of the model's performance. For medical application of ML models, besides metrics for the model performance on itself, the clinical effectiveness of the model should also be evaluated. However, this kind of evaluation can only be done after a model is developed. A potential approach is to use decision curve analysis, to quantify the benefit of using a model to guide actions. (165) Depending on the model task, different ML approaches are more likely to succeed. Choices can

be made for a supervised or unsupervised approach, for machine learning or deep learning algorithms and for specific models in each category. It is common practice to train several different types of models and then select the best performing ones for further development. (151)

Besides the subjects related to the model and the data, it is also important to thoroughly describe other conditions for successful development and implementation. Ethical and legal conditions, related to data privacy and METC clearance, should be investigated before accessing any data to prevent problems. For successful implementation, it is important that it is known what the requirements regarding information infrastructure in the hospital are and if they can be achieved. If this is not the case, this will prevent successful implementation. It is known that problems in technical integration are a major limiting factor for clinical AI implementation. (166) As is shown in figure 6.3, the acquisition and interpretation of images happens simultaneously in interventional cardiology. This means that an algorithm should be able process images and come to a conclusion in a short period of time. However, for algorithms used outside the cathlab this restraint is not applicable.

At this point it should also be assessed what the level of expertise of involved personnel is and if extra education is needed. This can consist of education on ML for physicians and providing medical background for technical professionals involved. As it is known that lack of knowledge on ML by physicians is also limiting implementation in clinical practice. (66) In Chapter 3, it became clear that a relatively thorough understanding of ML concepts by clinicians is needed not only for further acceptance and development of clinical ML models, but also when discussing possible applications in clinical practice. This is illustrated by the perceived potential value for ML in management of MINOCA, several participants stated that the potential value is low, as there are no clear angiographic features present in the ICA images of MINOCA patients. However, one might argue that DL models might be able to detect pattern that are not yet clear to operators at this time, as DL is able to find features in data that are not seen by humans or other algorithms. (126)

Most research on ML-based algorithms in healthcare is based on retrospective data (64) and it is likely that there is already some sort of repository or database present. So, it should be assessed what kind of data is present and if the data present is similar to the type of data described in the use case. Furthermore, it is essential to involve stakeholders at an early stage in the process, as developing a ML model for clinical practice requires knowledge from several domains. Relevant stakeholders in this process are clinicians, software engineers, data scientists, administrative leaders and clinical physicists as domain knowledge of medicine, software development, software implementation, data management, data acquisition and regulations is needed. (150) It might also be beneficial to involve patients in determining specific parts of the framework. (167) When creating the framework, it is possible that new insights acquired during the process, lead to adjustments in the descriptions of the previous steps, because of this it might be needed to iteratively go through the cycle several times to reach the best results.

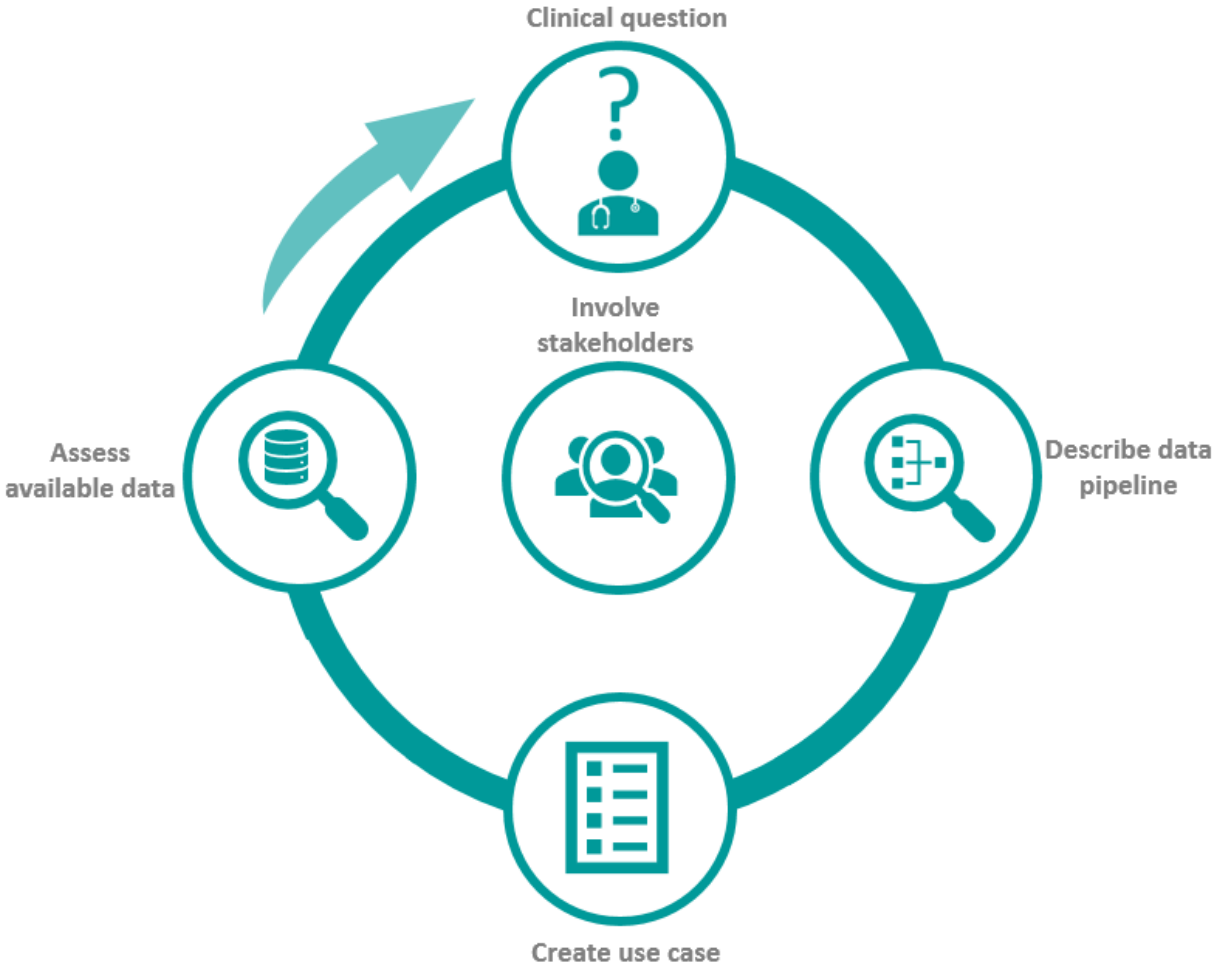


Fig. 6.4 The creation of a framework starts with defining the clinical question to be answered and describing the data and information pipeline. This information can be used to create a use case. After creating of the use case, it should be assessed if the available data is the data needed to solve the problem. New insights acquired while going through the creation of a framework can lead to adjustments of previous steps. Several iterations of the cycle might be needed to reach the best results. In every step relevant stakeholders should be consulted.

6.2.2 Dataset creation

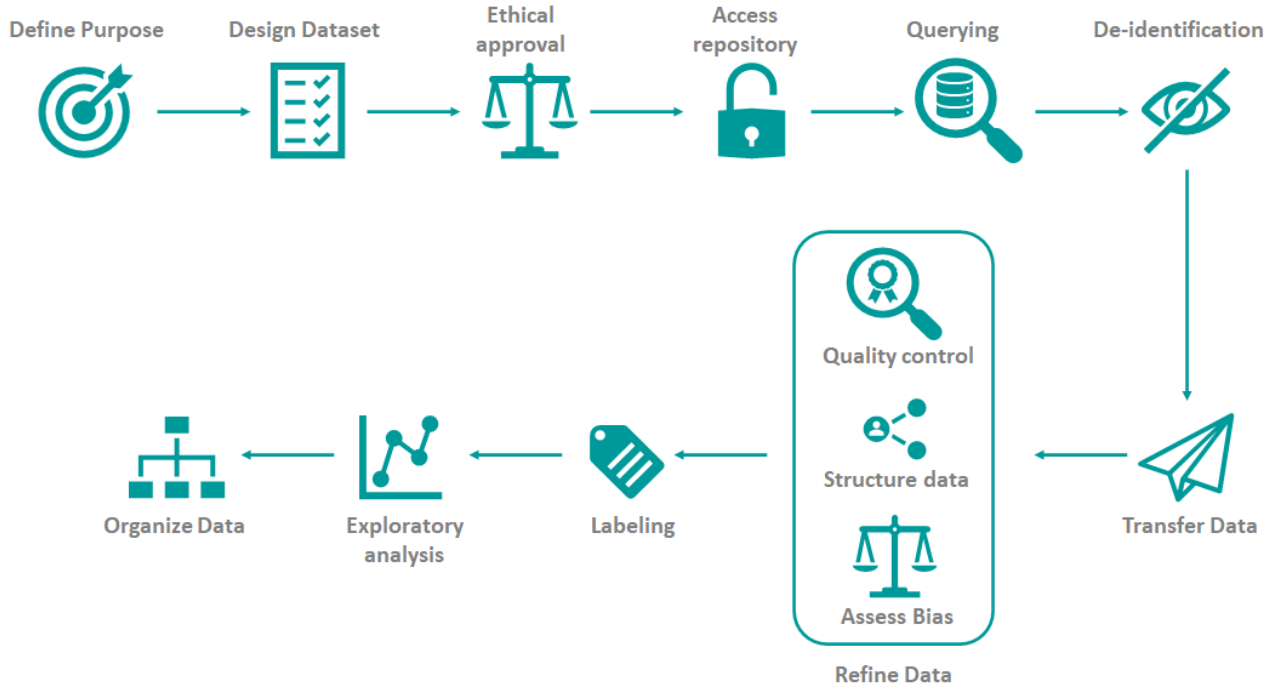


Fig. 6.5 A schematic overview for the creation of a dataset for machine learning.

The other major problem in creating ML model is the lack of high-quality annotated data sets that are representative for the real-world patient population. (64, 66, 150) The development of supervised ML models requires thoroughly curated data for optimal training, validation and testing. (64) At this moment, most parties involved in the development of AI models have limited access to medical images. (64) A small sample size leads to a lack of generalizability, bias and fitting on confounders, as this often results in a dataset that is not an accurate depiction of the real-world population. Besides the small sample size, inadequate curation and labeling of data also leads to models that do not perform well in clinical practice. There is an increasing number of open source data sets available, aiming to reduce the lack of appropriate data for ML algorithm development. (64) However, there are several limitations and drawback related to this type of datasets. First of all, there is not a data set available for every task. Most datasets available are related to radiology tasks and not to cardiology. (64) However, based on the systematic review in Chapter 2, there are several datasets available online containing ICA images. See tables 2.2-2.4 for the details. Privacy also poses a problem, due to privacy constraints not all metadata can be published online, decreasing the value of the dataset. (64) With open-source datasets there can be a wide variety in the quality of the data and often there is a lack of expert labeling or thorough curation. Due to these limitations, it can be necessary to create a dataset for the development of an ML model. However, this can be challenging. Therefore a systematic approach for data curation should be used.

Transfer learning

A commonly used technique to minimize the number of labeled cases needed for model training, is transfer learning. In transfer learning a model trained for one task is used as a basis for a new, related, task. It is especially popular in the development of deep learning models, as these types of models often require large amounts of data. (126) It is common to use transfer learning for the development of predictive models based on image or video data. Pre-trained deep learning models are accessible online for this purpose. These models are trained on a large number of images while prediction a large number of classes. In this way models are created that can efficiently extract features from imaging

data. Despite these images having different properties than images used in a medical context, using this as a starting point to train further has been proven to increase performance. (168) This process is illustrated in figure 6.6. It is important to carefully choose how much of the pre-trained model is used, as not all features are generic and more original dataset specific features are present in later layers of the model. (169) It should also be noted that transfer learning can only be applied to a new network with the same architecture as the network it will learn from.

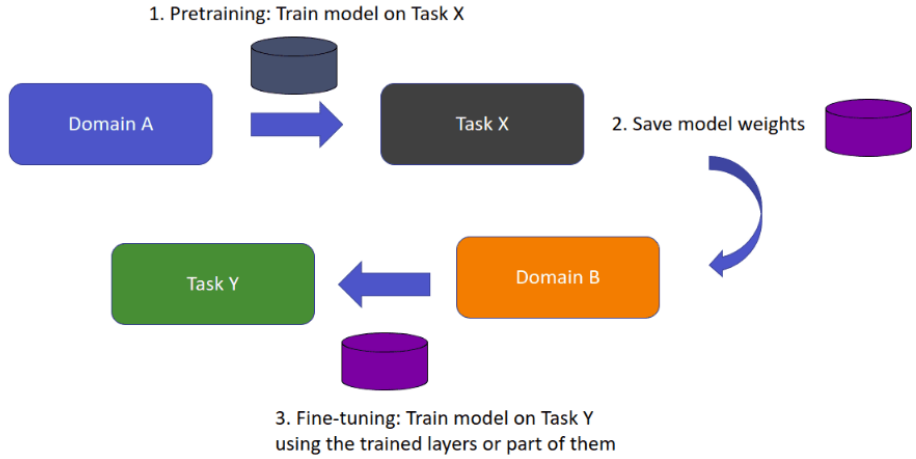


Fig. 6.6. A schematic representation of transfer learning. (170)

Purpose of the dataset

The creation of a dataset for the development of a ML model should always start with identifying the purpose of dataset and coupled to that the purpose of the ML model. (151) If the framework, described as above, is created the goal of the data set is in line with the use case. This includes the intent of the model, the inputs and the desired outputs. At this point it should also be assessed what factors could cause bias in the dataset. There are several factors that can lead to a biased data set. This can be factors regarding the population of included patients, such as age, sex, race and the distribution of diagnoses. (171) This can be seen as a mismatch between training data and real-world operational data. This mismatch can arise from deficiencies in the training set or from inappropriate application of the model, in an unintended patient context. (172) However, this bias is not always the result of a mismatch. With most medical conditions, patients with certain characteristics are more common than others. This will result in a large portion of the training data consisting of those type of patients. Due to this, applying the model to patients with atypical characteristics can cause problems. By conducting subgroup analyses with the model after training these effects can be measured and modified. (173) In the acquisition of the data bias can also arise. Factors of importance here are the used type of machines, acquisition protocols and differences in local clinical practice. This can lead to so called single-center or vendor bias. (64) The labeling process can also result in bias, especially when labels are manually created by experts or when a poorly defined ground truth is used. (174, 175)

In the acquisition of ICA images several sources of bias can be introduced, as the acquisition parameters are partially based on protocols and partially on operator preferences. There are protocols for which views should be used for each vessel, however, it is possible to vary the angle over a certain range within these views. Furthermore, X-ray settings can be varied. Choices can be made for the X-ray intensity, frames per second and the field of view used for acquisition. Besides this, there are various X-ray systems on the market and the used model can even vary between different cathlab rooms in one hospital. These factors should be considered when creating a dataset, as it can cause a bias.

Dataset design

After identifying the purpose, the data set should be designed. This starts with defining which data elements are needed. Then eligibility criteria for data records should be created. At this stage it should be determined if de-identification is needed and what data should be removed from each data record. Furthermore, quality criteria for the data records should be defined together with a protocol for handling missing data and outliers. It is important to involve clinicians in this step, as the absence or presence of specific data elements, such as lab values, can possibly provide information (e.g., if cardiac markers are assessed this might be an indication of the patients status). A strategy for determining labels for each data record should also be created. (151)

Ethical approval

During the design phase of the dataset, the project should be reviewed by a medical-ethical committee as patient data will be used. However, as most projects rely on retrospectively collected data and patients do not have to undergo additional procedures, explicit informed consent is often not required. (64) Besides this, ensuring data privacy and preventing any unethical decision making by the model can also be a part of the ethical clearance. As it is known that social inequalities can possibly be perturbated into the algorithm. (176)

Data Access

After laying out the intended design of the dataset, the next step is to access the medical data. The approach best suited depends on the parties involved and local infrastructure. When development is done in-house, it is often possible to directly access internal PACS or other data management systems. In the ideal situation a repository for the development of ML models is already present. Repository datasets are established to serve as the basis for datasets used to train, test, and validate ML Models. (151) However, most often this will not be the case. When an external party or in-house researchers without a treatment relationship with the patients are involved, data de-identification, in the form of pseudo-anonymization, must be done before granting access. The DICOM standard is the international standard to transmit, store, retrieve, print, process and display medical imaging data. (177) DICOM files contain, besides the pixel data, all sorts of sensitive information like patient names and identification numbers. This should be kept in mind while accessing medical imaging data.

Querying data

After the data storage system is accessible, queries can be created to extract relevant patients with the data elements relevant for model development. Strategy is dependent on local infrastructure and the dataset purpose. Queries can consists of strings, international classification of disease codes and procedural codes. (64) To further specify, it is also possible to apply the query on data acquired over a specific time period. As is shown in Chapter 4 and 5, there are several ways of querying EHR systems and image servers. However, scalability of the used approach is very case dependent. The querying strategies investigated in Chapter 4 show that the use of DBC-codes is not suitable for complex clinical entities such as MINOCA. However, this approach can give a coarse insight in the patient distribution. For complex cases as this, NLP can possibly be a viable direction. It was shown in a study by Redman et al. (2017) that the presence of fatty liver disease can accurately be detected by applying NLP on various imaging reports. (138) As described in Chapter 1, MINOCA can be identified by combining data from several modalities. The querying strategy applied in Chapter 5, for creating a dataset on functional relevance of stenoses with an objective ground truth, also needed a lot of manual, labor intensive, checking of querying results. However, various directions for improvement were identified. For these two subjects it became clear that additional automation techniques need to be developed to create a scalable strategy.

Data de-identification

After relevant patients and information are extracted from the available database via customized queries, sensitive information should be removed before continuing. According to the European General Data Protection Regulation, both retrospectively and prospectively gathered data require proper de-identification. (64) Medical images often have coupled DICOM metadata, which can hold sensitive information. There are tools available for automatic removal of this information. (178) It is also possible that patient information is embedded in images, which demands for additional procedures for de-identification. (179) It is also possible to anonymize the entire dataset by k-anonymity, which prevent potential hackers for determining patient identities. (180)

Data transfer

Whether data transfer is needed, depends on the local infrastructure and the parties involved in the model development. Possibilities are local storage and external storage at on on-premise server or storage in the cloud. (64)

Refining data records

After all data is accessible and possibly transferred to the right location, the data records should be refined. First, the data elements that are relevant for the model development should be selected. However, it is common to only select the relevant data elements while querying. In this way only the data elements of relevance for model development are collected. Next, all selected elements should be checked for quality, based on the predefined criteria. Depending on data types and sources it can be necessary to transform data elements. When data from multiple centers are used, the units of variables should be checked and transformed to the same unit. For images, resampling might be needed, as images acquired with different machines or settings can have different resolutions. The data should also be structured in a way that is convenient for model development. Data gaps and errors should be checked, missing values and outliers should be handled as described in the dataset design and duplicates should be removed. At this stage it is also good to reconsider the sources of bias determined at the start, to check if everything is covered. Relevant stakeholders should be included in this process, as specific domain knowledge can be needed to uncover all sources of bias.

The sources of bias in ICA images, as mentioned above, should be considered when refining data. Specifically, image size and the range of pixel values are important when preparing images for DL. These should be standardized in the dataset. (126) In Chapter 5 the choice was made to down sample all images to 512x512 pixels, to reduce memory usage and speed up training. The dataset contained both 8-bit and 16-bit images, so all images were set to an 8-bit representation. In Chapter 5 it also became clear that the used acquisition system and the used imaging protocols can vary, even within one center.

Labeling data records

After the data is checked for quality, the labeling strategy should be applied. The labels are the results that the ML model would produce given the input. Labeling strategy is dependent on the desired task. Most tasks regarding classification on medical images require more than the images as input for development. (122) Additional information based on further follow-up, other modalities or clinical outcomes can be used to create a ground truth. This process can be time and resource intensive, as accurate ground truth definition of a large number of data records is needed to build accurate models. (181) Structured reporting would strongly reduce time and costs for finding accurate ground truths. However, currently free-text reports are used in the majority of cases. (182) As most often retrospective data is used; labeling is also done retrospectively. This can be done in several ways, from

manual labeling to the application of natural-language processing models on reports and health records. (64)

An objective measure for determining data labels is desired, as a poorly defined ground truth can result in bias and decrease model performance. (174, 175) An example of an objective ground truth is the iFR measurements used in Chapter 5. In this case an objectively quantified value is used for labeling, this leaves no room for subjective interpretation. However, this is not possible for every dataset. For more complex ground truths, such as MINOCA in Chapter 4, the label is determined based on the interpretation of several sources of information. For these situation NLP might be a useful tool. As mentioned in Chapter 4, NLP can be used to extract information form EHRs and free-text reports with demonstrated success. (136-138)

Exploratory data analysis

After the data is cleaned, structured and labeled, the data can be explored further. By visualizing and transforming data, more information can be generated for further guidance of the model development. It also gives a chance to investigate the quality of the data and relations in and between features. This can be done by analyzing relations between features and labels and between features and other features. Results of the exploratory data analysis can indicate that the dataset or the research questions need more refining.

Organizing dataset

Next the data set should be split in a training, testing and validation group. The training dataset is used for the actual training of the model, the models learn how to handle the data based on this set. The validation set is used to evaluate the model fit on the training dataset, while optimizing model hyperparameters. The model sees the data, but it does not learn from it. The hyperparameters are adjusted based on the results of the model handling this set. The test set is used to evaluate the model after training and tuning of the hyperparameters is done. Often the validation set is used as test set, however, this does not give a fair evaluation of the model performance on real world data. The test set should contain a wide range of classes that the model would face when used on real world data.

Hyperparameter tuning

A ML model consists of parameters and hyperparameters. Examples of ML model parameters are the weights in artificial neural networks and the support vector in support vector machines. The parameters are the center of the ML model and are learned from the training data. The hyperparameters are used to control the learning process and model configurations. The amount of layers in a deep learning model is an example of a hyperparameter. (183) Hyperparameters are tuned to optimize the learning process. This can be done in various ways. Random search or grid search strategies are commonly used. In this process various values for the hyperparameters are evaluated and the optimal values are selected. (159)

Dataset splitting

Dataset splitting ratios depend mainly on the total amount of data available and the type of model that will be trained. Generally, datasets are split in 70:15:15 – 80:10:10 ratios for training, validation and testing. Testing with an external set, which is a dataset from another source than the training data, is preferred over an internal dataset as this can show the generalizability of the model. (184) There are several strategies for splitting the data. The dataset can be randomly split in the three different subsets. In classification problems with imbalanced datasets, it can be necessary to use a stratified split. In this way the ratio between labels in the subsets is kept equal to that in the whole dataset. (160)

It is also possible to split data based on certain conditions. In group based splitting all samples from one patient are either in the training or in the testing set. In time-based splitting all samples from before a specific moment are in the training set and all the samples after this moment in the testing set.

K-fold cross-validation

Another approach is k-fold cross-validation. This can be used when there is a small amount of data available, as it uses a limited sample for the estimation of model performance on unseen data. In k-fold cross-validation the dataset is randomly split into k groups. Each group is than alternately taken as testing set and the other groups as training. The model is than fitted on the training set and evaluated with the testing set. Then evaluation metrics are calculated for each group, which summarizes the model's performance. It is important to note that each sample is only used once for testing. (90) This method can also show signs of overfitting in the model. (185) However, k-fold cross-validation is a very computationally expensive method, as multiple iterations are necessary.

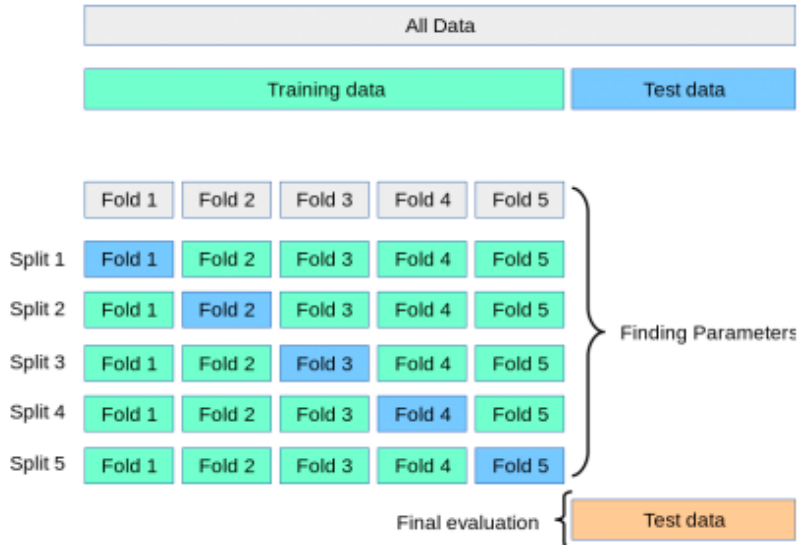


Fig. 6.7 Schematic representation of k-fold cross-validation. (186)

Assessing case balance

Typically, real world medical data is very imbalanced, which when not handled properly, can strongly decrease model performance. (187) Most ML models assume equal distributions and they are based on reaching the maximum overall classification accuracy. (188) This results in a model with a higher accuracy for the majority class and a low sensitivity for the minority class. (188) In most medical applications the minority class is the diseases or high risk population, so class imbalance can lead to errors with high-impact. (188) As described above, under framework, considering case-balance is important for selecting the right metrics. Selecting the right metrics can reduce problems coupled to case imbalance. There are several ways of handling imbalance. On data level, under- or over-sampling can be applied. In under-sampling data from the largest class is omitted and in over-sampling the amount of data from the minority class is increased by randomly sampling minority cases more than once. These methods are very straightforward, which makes them popular approaches. (188) However, this approach also has some drawbacks. With under-sampling, there is loss of information and with over-sampling model overfitting can occur. A more advanced technique, compared to random under- or oversampling, is to the synthetic minority oversampling technique (SMOTE). In SMOTE, samples in the minority class that are close to each other in feature space are selected. A line is drawn between the two samples and new samples are created on this line. (188) Combining under sampling techniques with oversampling is more effective than only under sampling the majority class.

(189) As mentioned in Chapter 5, creating synthetic data with even more sophisticated techniques, such as GAN, could also be used to increase dataset size or to balance case distributions.

It is also possible to address the imbalance problem on algorithm level, by biasing the classifier. A popular approach is cost-sensitive learning. In cost-sensitive learning a cost is coupled to a specific type of classification error, in medical context this would result in a higher cost for a high-impact mistake. Downside of this approach is that costs of the types of errors are not known in real world data and that the error rate in the true population can still exceed the determined limit, as real-world data is not equal to the data used for training. (188, 190)

6.3 Conclusion

The aim of this thesis was to investigate if it is possible to use ML to solve diagnostic challenges such as encountered in MINOCA and what is needed to apply ML on ICA images and other sources of coupled medical data. In this chapter it was attempted to answer these question by combining the results of subsequent chapters with scientific literature.

Expectations and perceived barriers on ML-based algorithms in clinical practice

The first research goal was to assess the expectations and perceived barriers by interventional cardiologists on ML based algorithms in clinical practice. Generally, the expectations for the future are positive and all participants expect a positive influence of ML on the clinical practice of interventional cardiology in the future. Furthermore, the willingness to collaborate in the development and clinical validation of ML algorithms is high. This is essential for translating ML models to clinical practice. It also became clear that education of clinicians is important for managing expectations and identifying barriers, as lack of trust in ML by clinicians was most frequently seen as a barriers.

Querying EHRs for the creation of a ML dataset

The second research goal of this thesis was to asses methods for querying EHRs based on health insurance codes, for the creation of a ML dataset. It became clear that this method was not feasible for creating datasets for complex clinical entities such as MINOCA. This approach resulted in a high amount of noise, due to which time-consuming and subjective manual analysis was needed. However, it was shown that the EHRs hold valuable data for the development of ML applications. These results can be seen as a starting point for further development of data collection strategies. An interesting and promising direction for further investigation is using NLP as method for querying EHRs in the creation of ML datasets.

Predicting lesion significance on ICA images

The third research goals was to conduct a proof-of-concept study on predicting coronary artery lesion significance on ICA images. In this study a strategy for dataset creation and the development of a DL network was created and tested. Similar to the querying methods investigated in chapter 4, the strategy for the dataset creation in this chapter required extensive manual analysis and inspection of the EHRs and image data. However, it was shown that the EHRs hold valuable information for the creation of DL networks and that it is feasible to train networks on this data. Furthermore, several research directions for automating collecting and processing EHR and image data were identified.

A roadmap for the curation of data for ML models in interventional cardiology

The last research goal was to create a roadmap for the curation of data for the development of ML models in interventional cardiology. In the development of ML models for clinical application, it is important to ensure that models are relevant for clinical practice and that the model is suited for the clinical workflow. By creating a framework describing the needed conditions for this, it is attempted to create a comprehensive overview of all relevant factors and to create a foundation for the development of a well performing and clinically applicable ML model. This can improve translation of ML models to clinical practice. Besides thoroughly considering the clinical practice, it is also important to create and use high-quality datasets for training, validation and testing. The described steps above aimed to give an overview of what is needed to create high-quality datasets, to reduce problems with generalizability, bias and confounders and perform well in a patient related context.

To conclude and answer the central questions of this thesis: ML shows promising results for solving diagnostic challenges in complex syndromes such as MINOCA. However, more research is needed to investigate the feasibility and performance of the identified directions. For the creation of algorithms that can be applied in clinical practice, close collaboration between ML professionals and clinicians is needed. Besides this, further research is needed to develop scalable strategies for the creation of large datasets, containing adequately labeled patients that represent the real life population.

7. References

1. Mendis S, Puska P, Norrving B, Organization WH. Global atlas on cardiovascular disease prevention and control: World Health Organization; 2011.
2. Sanchis-Gomar F, Perez-Quilis C, Leischik R, Lucia A. Epidemiology of coronary heart disease and acute coronary syndrome. Ann Transl Med. 2016;4(13):256-.
3. Lusis AJ. Atherosclerosis. Nature. 2000;407(6801):233-41.
4. Seenu R, Grub, KJ,. The Patient Guide to Heart, Lung, and Esophageal Surgery: A Website Presented by Cardiothoracic Surgeons Committed to Improving Patient Care. The Society of Thoracic Surgeons. April 2018. [cited 2021 april 12] Available from: https://ctsurgerypatients.org/adult-heart-disease/coronary-artery-disease [
5. Khan MA, Hashim MJ, Mustafa H, Baniyas MY, Al Suwaidi SKBM, AlKatheeri R, et al. Global Epidemiology of Ischemic Heart Disease: Results from the Global Burden of Disease Study. Cureus. 2020;12(7):e9349-e.
6. Mensah GA, Wei GS, Sorlie PD, Fine LJ, Rosenberg Y, Kaufmann PG, et al. Decline in Cardiovascular Mortality: Possible Causes and Implications. Circ Res. 2017;120(2):366-80.
7. Nivel Zorgregistraties eerste lijn, Nivel Zorgregistraties 2019. Avialable from: zorggevens.nl.
8. Altin C, Kanyilmaz S, Koc S, Gursoy YC, Bal U, Aydinalp A, et al. Coronary anatomy, anatomic variations and anomalies: a retrospective coronary angiography study. Singapore Med J. 2015;56(6):339-45.
9. Shriki JE SJ, Rashid MA, Hindoyan A, Withey JG, DeFrance A, Cunningham M, Oliveira GR, Warren BH, Wilcox A. Identifying, characterizing, and classifying congenital anomalies of the coronary arteries. (1527-1323 (Electronic)).
10. Angelini P. Coronary artery anomalies--current clinical issues: definitions, classification, incidence, clinical relevance, and treatment guidelines. (0730-2347 (Print)).
11. Greenwood JP, Ripley DP, Berry C, McCann GP, Plein S, Bucciarelli-Ducci C, et al. Effect of Care Guided by Cardiovascular Magnetic Resonance, Myocardial Perfusion Scintigraphy, or NICE Guidelines on Subsequent Unnecessary Angiography Rates: The CE-MARC 2 Randomized Clinical Trial. (1538-3598 (Electronic)).
12. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. (0028-4793 (Print)).
13. De Bruyne B, Pijls NH, Kalesan B, Barbato E, Tonino PA, Piroth Z, et al. Fractional flow reserve-guided PCI versus medical therapy in stable coronary disease. (1533-4406 (Electronic)).
14. Tonino P, A,, B DB, Pijls NH, Siebert U, Ikeno F, van' t Veer M, et al. Fractional flow reserve versus angiography for guiding percutaneous coronary intervention. (1533-4406 (Electronic)).
15. Knuuti J, Ballo H, Juarez-Orozco LE, Saraste A, Kolh P, Rutjes AWS, et al. The performance of non-invasive tests to rule-in and rule-out significant coronary artery stenosis in patients with stable angina: a meta-analysis focused on post-test disease probability. European Heart Journal. 2018;39(35):3322-30.
16. Tonino P, A,, Fearon WF, De Bruyne B, Oldroyd KG, Leesar MA, Ver Lee PN, et al. Angiographic versus functional severity of coronary artery stenoses in the FAME study fractional flow reserve versus angiography in multivessel evaluation. (1558-3597 (Electronic)).
17. Neumann FJ, Sousa-Uva M, Ahlsson A, Alfonso F, Banning AP, Benedetto U, et al. 2018 ESC/EACTS Guidelines on myocardial revascularization. (1522-9645 (Electronic)).
18. Escaned J, Echavarría-Pinto M, Garcia-Garcia HM, van de Hoef TP, de Vries T, Kaul P, et al. Prospective Assessment of the Diagnostic Accuracy of Instantaneous Wave-Free Ratio to Assess Coronary Stenosis Relevance: Results of ADVISE II International, Multicenter Study (ADenosine Vasodilator Independent Stenosis Evaluation II). (1876-7605 (Electronic)).
19. Davies JE, Sen S, Dehbi HM, Al-Lamee R, Petraco R, Nijjer SS, et al. Use of the Instantaneous Wave-free Ratio or Fractional Flow Reserve in PCI. (1533-4406 (Electronic)).

20. Petraco R, Al-Lamee R, Gotberg M, Sharp A, Hellig F, Nijjer SS, et al. Real-time use of instantaneous wave-free ratio: results of the ADVISE in-practice: an international, multicenter evaluation of instantaneous wave-free ratio in clinical practice. Am Heart J. 2014;168(5):739-48.
21. Tebaldi M, Campo G, Biscaglia S. Fractional flow reserve: Current applications and overview of the available data. World J Clin Cases. 2015;3(8):678-81.
22. Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes: The Task Force for the diagnosis and management of chronic coronary syndromes of the European Society of Cardiology (ESC). European Heart Journal. 2020;41(3):407-77.
23. Oro R. Percutaneous coronary intervention (PCI). 2019. [Cited on 2021 april 13] Available from: https://rodnieoro.com/home/percutaneous-coronary-intervention-angioplasty.
24. Kok JN, Boers EJ, Kosters WA, Van der Putten P, Poel MJAi. Artificial intelligence: definition, trends, techniques, and cases. 2009;1:270-99.
25. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Family Med Prim Care. 2019;8(7):2328-31.
26. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Medical Research Methodology. 2019;19(1):64.
27. Nichols JA, Herbert Chan HW, Baker MAB. Machine learning: applications of artificial intelligence to imaging and diagnosis. Biophys Rev. 2019;11(1):111-8.
28. Chen XW, Gao Jean X. Big Data Bioinformatics. Methods. 2016;111:1-2.
29. Paradzhanyan L. Demystifying Deep Learning and Artificial Intelligence [Internet]. The New Stack; May 8 2020 [Cited April 14 2021] Available from: https://thenewstack.io/demystifying-deep-learning-and-artificial-intelligence/.
30. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nature Medicine. 2019;25(1):24-9.
31. Krizhevsky A SI, Hinton GE. Imagenet classification with deep convolutional neural networks. In Adv Neural Inf Process Syst. 2012:1097-105.
32. He K, Zhang X, Ren S, Sun J, editors. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015 IEEE International Conference on Computer Vision (ICCV); 2015 7-13 Dec. 2015.
33. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. Artificial Intelligence in Healthcare. 2020:25-60.
34. Mintz Gary S, Popma Jeffrey J, Pichard Augusto D, Kent Kenneth M, Satler Lowell F, Chien Chuang Y, et al. Limitations of Angiography in the Assessment of Plaque Distribution in Coronary Artery Disease. Circulation. 1996;93(5):924-31.
35. Miller SaBL. Ischemic heart disease [Internet]. Radiologykey; Dec 26 2016. [Cited on April 15 2021]. Available from: https://radiologykey.com/ischemic-heart-disease/.
36. Occleshaw JG, SC. and Gerber IL. Cardiac testing [Internet]. Clinical gate; Feb 13 2015. [Cited on April 14 2021]. Available from: https://clinicalgate.com/cardiac-testing/.
37. Zhang H, Mu L, Hu S, Nallamothu BK, Lansky AJ, Xu B, et al. Comparison of Physician Visual Assessment With Quantitative Coronary Angiography in Assessment of Stenosis Severity in China. JAMA Intern Med. 2018;178(2):239-47.
38. Nogic J, Prosser H, O'Brien J, Thakur U, Soon K, Proimos G, et al. The assessment of intermediate coronary lesions using intracoronary imaging. Cardiovasc Diagn Ther. 2020;10(5):1445-60.
39. Koskinas K, C,, Nakamura M, Räber L, Colleran R, Kadota K, Capodanno D, et al. Current use of intracoronary imaging in interventional practice - Results of a European Association of Percutaneous Cardiovascular Interventions (EAPCI) and Japanese Association of Cardiovascular Interventions and Therapeutics (CVIT) Clinical Practice Survey. (1969-6213 (Electronic)).
40. Koganti S, Kotecha T, Rakhit RD. Choice of Intracoronary Imaging: When to use Intravascular Ultrasound or Optical Coherence Tomography. Interv Cardiol. 2016;11(1):11-6.

41. Herscovici R, Sedlak T, Wei J, Pepine CJ, Handberg E, Bairey Merz CN. Ischemia and No Obstructive Coronary Artery Disease (INOCA): What Is the Risk? J Am Heart Assoc. 2018;7(17):e008868-e.
42. Jespersen L, Hvelplund A, Abildstrøm S, Pedersen F, Galatius S, Madsen J, et al. Stable angina pectoris with no obstructive coronary artery disease is associated with increased risks of major adverse cardiovascular events. (1522-9645 (Electronic)).
43. Bairey Merz CN, Pepine Carl J, Walsh Mary N, Fleg Jerome L, Camici Paolo G, Chilian William M, et al. Ischemia and No Obstructive Coronary Artery Disease (INOCA). Circulation. 2017;135(11):1075-92.
44. Jespersen L, Abildstrom SZ, Hvelplund A, Madsen JK, Galatius S, Pedersen F, et al. Burden of hospital admission and repeat angiography in angina pectoris patients with and without coronary artery disease: a registry-based cohort study. PLoS One. 2014;9(4):e93170-e.
45. Kunadian V, Chieffo A, Camici PG, Berry C, Escaned J, Maas AHEM, et al. An EAPCI Expert Consensus Document on Ischaemia with Non-Obstructive Coronary Arteries in Collaboration with European Society of Cardiology Working Group on Coronary Pathophysiology & Microcirculation Endorsed by Coronary Vasomotor Disorders International Study Group. European Heart Journal. 2020;41(37):3504-20.
46. Galassi A, Crea F, Araujo L, Lammertsma A, Pupita G, Yamamoto Y, et al. Comparison of regional myocardial blood flow in syndrome X and one-vessel coronary artery disease. (0002-9149 (Print)).
47. Panting J, Gatehouse P, Yang G, Grothues F, Firmin D, Collins P, et al. Abnormal subendocardial perfusion in cardiac syndrome X detected by cardiovascular magnetic resonance imaging. (1533-4406 (Electronic)).
48. Mejía-Rentería H, van der Hoeven N, van de Hoef TP, Heemelaar J, Ryan N, Lerman A, et al. Targeting the dominant mechanism of coronary microvascular dysfunction with intracoronary physiology tests. (1875-8312 (Electronic)).
49. Pries AR, Badimon L, Bugiardini R, Camici PG, Dorobantu M, Duncker DJ, et al. Coronary vascular regulation, remodelling, and collateralization: mechanisms and clinical implications on behalf of the working group on coronary pathophysiology and microcirculation. (1522-9645 (Electronic)).
50. Sorop O, Merkus D, de Beer V, Houweling B, Pistea A, McFalls EO, et al. Functional and structural adaptations of coronary microvessels distal to a chronic coronary artery stenosis. (1524-4571 (Electronic)).
51. Sorop O, van den Heuvel M, van Ditzhuijzen NS, de Beer VJ, Heinonen I, van Duin RW, et al. Coronary microvascular dysfunction after long-term diabetes and hypercholesterolemia. (1522-1539 (Electronic)).
52. Beltrame JF, Crea F, Kaski JC, Ogawa H, Ong P, Sechtem U, et al. International standardization of diagnostic criteria for vasospastic angina. (1522-9645 (Electronic)).
53. Ong P, Athanasiadis A, Perne A, Mahrholdt H, T S, Hill S, et al. Coronary vasomotor abnormalities in patients with stable angina after successful stent implantation but without in-stent restenosis. (1861-0692 (Electronic)).
54. Everaars H, de Waard GA, Driessen RS, Danad I, van de Ven PM, Raijmakers PG, et al. Doppler Flow Velocity and Thermodilution to Assess Coronary Flow Reserve: A Head-to-Head Comparison With [(15)O]H(2)O PET. (1876-7605 (Electronic)).
55. Barbato E AW, Aengevaeren WR, Werner G, Klauss V, Bojara W, Herzfeld I, Oldroyd KG, Pijls NH, De Bruyne, B,. Validation of coronary flow reserve measurements by thermodilution in clinical practice. (0195-668X (Print)).
56. Fearon WF, Balsam LB, Farouque HM, Caffarelli AD, Robbins RC, Fitzgerald PJ, et al. Novel index for invasively assessing the coronary microcirculation. (1524-4539 (Electronic)).
57. Ong P, Athanasiadis A, Sechtem U. Intracoronary Acetylcholine Provocation Testing for Assessment of Coronary Vasomotor Disorders. LID - 10.3791/54295 [doi] LID - 54295. (1940-087X (Electronic)).

58. Pasupathy S, Air T, Dreyer RP, Tavella R, Beltrame JF. Systematic review of patients presenting with suspected myocardial infarction and nonobstructive coronary arteries. (1524-4539 (Electronic)).
59. Pustjens TFS, Appelman Y, Damman P, ten Berg JM, Jukema JW, de Winter RJ, et al. Guidelines for the management of myocardial infarction/injury with non-obstructive coronary arteries (MINOCA): a position paper from the Dutch ACS working group. Netherlands Heart Journal. 2020;28(3):116-30.
60. Agewall S, Beltrame JF, Reynolds HR, Niessner A, Rosano G, Caforio ALP, et al. ESC working group position paper on myocardial infarction with non-obstructive coronary arteries. European Heart Journal. 2017;38(3):143-53.
61. Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA, et al. Fourth universal definition of myocardial infarction (2018). European Heart Journal. 2019;40(3):237-69.
62. Siegersma KR, Leiner T, Chew DP, Appelman Y, Hofstra L, Verjans JW. Artificial intelligence in cardiovascular imaging: state of the art and implications for the imaging cardiologist. Neth Heart J. 2019;27(9):403-13.
63. Betancur J, Commandeur F, Motlagh M, Sharir T, Einstein AJ, Bokhari S, et al. Deep Learning for Prediction of Obstructive Disease From Fast Myocardial Perfusion SPECT: A Multicenter Study. JACC Cardiovasc Imaging. 2018;11(11):1654-63.
64. Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, et al. Preparing Medical Imaging Data for Machine Learning. Radiology. 2020;295(1):4-15.
65. Shah NH, Milstein A, Bagley PSC. Making Machine Learning Models Clinically Useful. JAMA. 2019;322(14):1351-2.
66. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine. 2019;17(1):195.
67. Konst RE, Meeder JG, Wittekoek ME, Maas AHEM, Appelman Y, Piek JJ, et al. Ischaemia with no obstructive coronary arteries. Neth Heart J. 2020;28(Suppl 1):66-72.
68. Nallamothu B, K,, Spertus JA, Lansky AJ, Cohen DJ, Jones PG, Kureshi F, et al. Comparison of clinical interpretation with visual assessment and quantitative coronary angiography in patients undergoing percutaneous coronary intervention in contemporary practice: the Assessing Angiography (A2) project. (1524-4539 (Electronic)).
69. Kwak S, Lee Y, Ko T, Yang S, Hwang IC, Park JB, et al. Unsupervised Cluster Analysis of Patients With Aortic Stenosis Reveals Distinct Population With Different Phenotypes and Outcomes. (1942-0080 (Electronic)).
70. Mazurowski MA, Buda M, Saha A, Bashir MR. Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. J Magn Reson Imaging. 2019;49(4):939-54.
71. Wang L, Liang D-x, Yin X-l, Qiu J, Yang Z-y, Xing J-h, et al. Coronary Artery Segmentation in Angiographic Videos Using A 3D-2D CE-Net. arXiv e-prints. 2020:arXiv:2003.11851.
72. Au B, Shaham U, Dhruva S, Bouras G, Cristea E, Lansky A, et al. Automated Characterization of Stenosis in Invasive Coronary Angiography Images with Convolutional Neural Networks. 2018;abs/1807.10597.
73. Hernandez-Vela A, Gatta C, Escalera S, Igual L, Martin-Yuste V, Sabate M, et al. Accurate Coronary Centerline Extraction, Caliber Estimation, and Catheter Detection in Angiographies. IEEE Transactions on Information Technology in Biomedicine. 2012;16(6):1332-40.
74. Fan J, Yang J, Wang Y, Yang S, Ai D, Huang Y, et al. Multichannel Fully Convolutional Network for Coronary Artery Segmentation in X-Ray Angiograms. IEEE Access. 2018;6:44635-43.
75. Ciusdel C, Turcea A, Puiu A, Itu L, Calmac L, Weiss E, et al. Deep Neural Networks for ECG-free Cardiac Phase and End-Diastolic Frame Detection on Coronary Angiographies. 2020;84:101749.
76. Fan J, Yang J, Wang Y, Yang S, Ai D, Huang Y, et al. Deep feature descriptor based hierarchical dense matching for X-ray angiographic images. Computer Methods and Programs in Biomedicine. 2019;175:233-42.

77. Ma H, Ambrosini P, Walsum TV, editors. Fast Prospective Detection of Contrast Inflow in X-ray Angiograms with Convolutional Neural Network and Recurrent Neural Network. MICCAI; 2017.
78. Fang H, Yang J, Zhu J, Ai D, Huang Y, Jiang Y, et al. Greedy Graph Searching for Vascular Tracking in Angiographic Image Sequences. 2018;abs/1805.09940.
79. Ripley BD, Hjort NL. Pattern Recognition and Neural Networks: Cambridge University Press; 1995.
80. Yang S, Kweon J, Roh J-H, Lee J-H, Kang H, Park L-J, et al. Deep learning segmentation of major vessels in X-ray coronary angiography. Scientific Reports. 2019;9(1):16897.
81. Yang S, Yang J, Wang Y, Yang Q, Ai D, Wang Y. Automatic Coronary Artery Segmentation in X-ray Angiograms by Multiple Convolutional Neural Networks. Proceedings of the 3rd International Conference on Multimedia and Image Processing; Guiyang, China: Association for Computing Machinery; 2018. p. 31–5.
82. Liu X, Yang R, Xie L, Zhang H, Xu B. TCT-242 Detection and Classification of Coronary Bifurcation Lesions by Using Artificial Intelligence. Journal of the American College of Cardiology. 2019;74(13_Supplement):B241-B.
83. Liu X, Du T, Zhang H, Sun C, editors. Detection and Classification of Chronic Total Occlusion lesions using Deep Learning. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2019 23-27 July 2019.
84. Yang S, Kweon J, Kim Y-H, Roh J-H, Kang D-Y, Lee Pil H, et al. TCT-194 A Fully Automated Classification and Segmentation of X-Ray Coronary Angiography Using Deep Learning Approach. Journal of the American College of Cardiology. 2019;74(13_Supplement):B193-B.
85. Litjens G, Ciompi F, Wolterink JM, de Vos BD, Leiner T, Teuwen J, et al. State-of-the-Art Deep Learning in Cardiovascular Image Analysis. JACC: Cardiovascular Imaging. 2019;12(8, Part 1):1549-65.
86. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A Survey of Methods for Explaining Black Box Models. 2018;51(5 %J ACM Comput. Surv.):Article 93.
87. Lapuschkin S, Wäldchen S, Binder A, Montavon G, Samek W, Müller K-R. Unmasking Clever Hans predictors and assessing what machines really learn. Nat Commun. 2019;10(1):1096-.
88. Hae HA-O, Kang SA-O, Kim WA-O, Choi SY, Lee JA-O, Bae Y, et al. Machine learning assessment of myocardial ischemia using angiography: Development and retrospective validation. (1549-1676 (Electronic)).
89. Cho H, Lee JG, Kang SJ, Kim WJ, Choi SY, Ko J, et al. Angiography-Based Machine Learning for Predicting Fractional Flow Reserve in Intermediate Coronary Artery Lesions. J Am Heart Assoc. 2019;8(4):e011685.
90. Gareth James DW, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning. 1 ed. New York: Springer; 2013. 426 p.
91. Cruz-Aceves I, Oloumi F, Rangayyan RM, Aviña-Cervantes JG, Hernandez-Aguirre A. Automatic segmentation of coronary arteries using Gabor filters and thresholding based on multiobjective optimization. Biomedical Signal Processing and Control. 2016;25:76-85.
92. Qin B, Jin M, Hao D, Lv Y, Liu Q, Zhu Y, et al. Accurate vessel extraction via tensor completion of background layer in X-ray coronary angiograms. Pattern Recognition. 2019;87:38-54.
93. Alizadehsani R, Roshanzamir M, Abdar M, Beykikhoshk A, Khosravi A, Panahiazar M, et al. A database for using machine learning and data mining techniques for coronary artery disease diagnosis. Sci Data. 2019;6(1):227-.
94. Kern MJ, Samady H. Current concepts of integrated coronary physiology in the catheterization laboratory. (1558-3597 (Electronic)).
95. Balanescu S. Fractional Flow Reserve Assessment of Coronary Artery Stenosis. Eur Cardiol. 2016;11(2):77-82.
96. Hao D, Ding S, Qiu L, Lv Y, Fei B, Zhu Y, et al. Sequential vessel segmentation via deep channel attention network. Neural Networks. 2020;128:172-87.
97. Piayda K, Kleinebrecht L, Afzal S, Bullens R, Ter Horst I, Polzin A, et al. Dynamic coronary roadmapping during percutaneous coronary intervention: a feasibility study. Eur J Med Res. 2018;23(1):36-.

98. Patel M, R,, Peterson E, Dai D, Brennan JM, Redberg RF, Anderson HV, et al. Low diagnostic yield of elective coronary angiography. (1533-4406 (Electronic)).
99. Pepine CJ, Ferdinand KC, Shaw LJ, Light-McGroary KA, Shah RU, Gulati M, et al. Emergence of Nonobstructive Coronary Artery Disease: A Woman's Problem and Need for Change in Definition on Angiography. Journal of the American College of Cardiology. 2015;66(17):1918-33.
100. Yu F, Zhao J, Gong Y, Wang Z, Li Y, Yang F, et al., editors. Annotation-Free Cardiac Vessel Segmentation via Knowledge Transfer from Retinal Images. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019; 2019 2019//; Cham: Springer International Publishing.
101. Zhao L, Li D, Chen J, Wan T, editors. Automated Coronary Tree Segmentation for X-ray Angiography Sequences Using Fully-convolutional Neural Networks. 2018 IEEE Visual Communications and Image Processing (VCIP); 2018 9-12 Dec. 2018.
102. Cervantes-Sanchez F C-AI, Hernandez-Aguirre A, Hernandez-Gonzalez MA, Solorio-Meza SE. Automatic Segmentation of Coronary Arteries in X-ray Angiograms using Multiscale Analysis and Artificial Neural Networks. Applied Sciences. 2019;9(24):5507.
103. Ma B, Liu S, Zhi Y, Song Q. Flow Based Self-supervised Pixel Embedding for Image Segmentation. arXiv e-prints. 2019:arXiv:1901.00520.
104. Karapataki M, De Wilde P. Hopfield network applied to blood vessel detection in angiograms. Medical and Biological Engineering and Computing. 1997;35(4):428-30.
105. Dongdong H, Yiming L, Binjie Q, editors. Learning saliently temporal-spatial features for x-ray coronary angiography sequence segmentation. ProcSPIE; 2019.
106. Yang S, Kweon J, Kim Y, editors. Major Vessel Segmentation on X-ray Coronary Angiography using Deep Networks with a Novel Penalty Loss Function2019.
107. Plourde M, Luc D, editors. Multi scale classification approach for coronary artery detection from X-ray angiography. 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA); 2012 2-5 July 2012.
108. Cruz-Aceves IA-O, Cervantes-Sanchez F, Avila-Garcia MA-O. A Novel Multiscale Gaussian-Matched Filter Using Neural Networks for the Segmentation of X-Ray Coronary Angiograms. (2040-2295 (Print)).
109. Jo K, Kweon J, Kim Y, Choi J. Segmentation of the Main Vessel of the Left Anterior Descending Artery Using Selective Feature Mapping in Coronary Angiography. IEEE Access. 2019;7:919-30.
110. Nasr-Esfahani E, Karimi N, Jafari MH, Soroushmehr SMR, Samavi S, Nallamothu BK, et al. Segmentation of vessels in angiograms using convolutional neural networks. Biomedical Signal Processing and Control. 2018;40:240-51.
111. Jun TJ, Kweon J, Kim Y-H, Kim D. T-Net: Nested encoder–decoder architecture for the main vessel segmentation in coronary angiography. Neural Networks. 2020;128:216-33.
112. Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SM, Ward K, Jafari MH, et al. Vessel extraction in X-ray angiograms using deep learning. (2694-0604 (Electronic)).
113. Kumar S, Amutha RJSR, Essays. Automated classification of coronary artery disease using discrete wavelet transform and back propagation neural network. 2014;9:440-51.
114. Cong C, Kato Y, Vasconcellos HD, Lima J, Venkatesh B, editors. Automated Stenosis Detection and Classification in X-ray Angiography Using Deep Neural Network. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2019 18-21 Nov. 2019.
115. Du T, Xie L, Zhang H, Liu X, Wang X, Chen D, et al. Automatic and multimodal analysis for coronary angiography: training and validation of a deep learning architecture. LID - EIJ-D-20-00570 [pii] LID - 10.4244/EIJ-D-20-00570 [doi]. (1969-6213 (Electronic)).
116. Wu W, Zhang J, Xie H, Zhao Y, Zhang S, Gu L. Automatic detection of coronary artery stenosis by convolutional neural network with temporal constraint. Computers in Biology and Medicine. 2020;118:103657.
117. Xavier L, Michal N, Mahmoudi S, Saïd M, Piussi C. Deep Convolutional Neural Networks (CNN) for classification of coronary arteries. International Journal of Computer Assisted Radiology and Surgery. 2020.

118. Chen S, Tang Y, Shi X, Zhang H, Xie L, Xu B, editors. Convolution Pyramid Network: A Classification Network on Coronary Artery Angiogram Images. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2020 20-24 July 2020.
119. Lee Paul C, Lee N, Pyo R. Abstract 12950: Convolutional Neural Networks for Interpretation of Coronary Angiography. Circulation. 2019;140(Suppl_1):A12950-A.
120. Zhang D, Yang G, Zhao S, Zhang Y, Zhang H, Li SJA. Direct Quantification for Coronary Artery Stenosis Using Multiview Learning. 2019;abs/1907.10032.
121. Du T, Liu X, Zhang H, Xu B, editors. Real-time Lesion Detection of Cardiac Coronary Artery Using Deep Neural Networks. 2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC); 2018 22-24 Aug. 2018.
122. Allen B, Jr., Seltzer SE, Langlotz CP, Dreyer KP, Summers RM, Petrick N, et al. A Road Map for Translational Research on Artificial Intelligence in Medical Imaging: From the 2018 National Institutes of Health/RSNA/ACR/The Academy Workshop. (1558-349X (Electronic)).
123. Siau K, Wang W. Trusting Artificial Intelligence in Healthcare2018.
124. Huisman M, Ranschaert E, Parker W, Mastrodicasa D, Koci M, Pinto de Santos D, et al. An international survey on AI in radiology in 1,041 radiologists and radiology residents part 1: fear of replacement, knowledge, and attitude. European Radiology. 2021.
125. Hartnell GG, Parnell BM, Pridie RB. Coronary artery ectasia. Its prevalence and clinical significance in 4993 patients. (0007-0769 (Print)).
126. Goodfellow I, Bengio Y, Courville A. Deep Learning: The MIT Press; 2016.
127. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3(1):160018.
128. Lin W-C, Chen JS, Chiang MF, Hribar MR. Applications of Artificial Intelligence to Electronic Health Record Data in Ophthalmology. Transl Vis Sci Technol. 2020;9(2):13-.
129. Evans RS. Electronic Health Records: Then, Now, and in the Future. Yearb Med Inform. 2016;Suppl 1(Suppl 1):S48-S61.
130. Norton PT, Rodriguez HP, Shortell SM, Lewis VA. Organizational influences on healthcare system adoption and use of advanced health information technology capabilities. Am J Manag Care. 2019;25(1):e21-e5.
131. Eindhoven DC, Hilt AD, Zwaan TC, Schalij MJ, Borleffs CJW. Age and gender differences in medical adherence after myocardial infarction: Women do not receive optimal treatment – The Netherlands claims database. European Journal of Preventive Cardiology. 2018;25(2):181-9.
132. Eindhoven DC, van Staveren LN, van Erkelens JA, Ikkersheim DE, Cannegieter SC, Umans VAWM, et al. Nationwide claims data validated for quality assessments in acute myocardial infarction in the Netherlands. Neth Heart J. 2018;26(1):13-20.
133. Eindhoven DC, Wu HW, Kremer SWF, van Erkelens JA, Cannegieter SC, Schalij MJ, et al. Mortality differences in acute myocardial infarction patients in the Netherlands: The weekend-effect. Am Heart J. 2018;205:70-6.
134. van der Velde E, Atsma DE, Schalij M, Witteman TA, Foeken H, Bruijn FDB. Development and implementation of a fully paperless cardiology information system (EPD-Vision)2006. 849-52 p.
135. Sachdev E, Bairey Merz CN, Mehta PK. Takotsubo Cardiomyopathy. Eur Cardiol. 2015;10(1):25-30.
136. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review. JMIR Med Inform. 2019;7(2):e12239.
137. Van Vleck TT, Chan L, Coca SG, Craven CK, Do R, Ellis SB, et al. Augmented intelligence with natural language processing applied to electronic health records for identifying patients with non-alcoholic fatty liver disease at risk for disease progression. International Journal of Medical Informatics. 2019;129:334-41.
138. Redman JS, Natarajan Y, Hou JK, Wang J, Hanif M, Feng H, et al. Accurate Identification of Fatty Liver Disease in Data Warehouse Utilizing Natural Language Processing. Digestive Diseases and Sciences. 2017;62(10):2713-8.

139. Suzuki N, Asano T, Nakazawa G, Aoki J, Tanabe K, Hibi K, et al. Clinical expert consensus document on quantitative coronary angiography from the Japanese Association of Cardiovascular Intervention and Therapeutics. Cardiovasc Interv Ther. 2020;35(2):105-16.
140. Chung W-Y, Choi B-J, Lim S-H, Matsuo Y, Lennon RJ, Gulati R, et al. Three dimensional quantitative coronary angiography can detect reliably ischemic coronary lesions based on fractional flow reserve. J Korean Med Sci. 2015;30(6):716-24.
141. Sejr-Hansen M, Westra J, Winther S, Tu S, Nissen L, Gormsen L, et al. Comparison of quantitative flow ratio and fractional flow reserve with myocardial perfusion scintigraphy and cardiovascular magnetic resonance as reference standard. A Dan-NICAD substudy. The International Journal of Cardiovascular Imaging. 2020;36(3):395-402.
142. Peper J, van Hamersvelt RW, Rensing BJWM, van Kuijk J-P, Voskuil M, Berg JMt, et al. Diagnostic performance and clinical implications for enhancing a hybrid quantitative flow ratio–FFR revascularization decision-making strategy. Scientific Reports. 2021;11(1):6425.
143. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging. 2018;9(4):611-29.
144. Simonyan K, Zisserman AJapa. Very deep convolutional networks for large-scale image recognition. 2014.
145. Garbin C, Zhu X, Marques O. Dropout vs. batch normalization: an empirical study of their impact to deep learning. Multimedia Tools and Applications. 2020;79(19):12777-815.
146. Hussain Z, Gimenez F, Yi D, Rubin D. Differential Data Augmentation Techniques for Medical Imaging Classification Tasks. AMIA Annu Symp Proc. 2018;2017:979-84.
147. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D, editors. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE international conference on computer vision; 2017.
148. Lan L, You L, Zhang Z, Fan Z, Zhao W, Zeng N, et al. Generative Adversarial Networks and Its Applications in Biomedical Informatics. Front Public Health. 2020;8:164-.
149. Huang S-C, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. npj Digital Medicine. 2020;3(1):136.
150. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. Nature Medicine. 2019;25(9):1337-40.
151. Genereaux B, Bialecki B, Diedrich K, O'Donnell K, Roth C, Schroeder A, et al. IHE Radiology Whitepaper, AI interoperability in Imaging. 2021.
152. Tirado-Ramos A, Hu J, Lee K. Information object definition-based unified modeling language representation of DICOM structured reporting: a case study of transcoding DICOM to XML. (1067-5027 (Print)).
153. Safdari R, Farzi J, Ghazisaeidi M, Mirzaee M, Goodini A. The application of use case modeling in designing medical imaging information systems. ISRN Radiol. 2013;2013:530729-.
154. Kohli MD, Summers RM, Geis JR. Medical Image Data and Datasets in the Era of Machine Learning-Whitepaper from the 2016 C-MIMI Meeting Dataset Session. (1618-727X (Electronic)).
155. JI. W, Willemink MJ, De Cecco CN. Artificial Intelligence and Machine Learning in Radiology: Current State and Considerations for Routine Clinical Implementation. (1536-0210 (Electronic)).
156. Parks J. How to Identify Machine Learning Use Cases in Your Business [Internet]. Agile Thought; 2021. [Cited on April 10 2021] Available on: https://agilethought.com/blogs/how-to-identify-machine-learning-business-use-case/ 2021 [
157. Santeramo R, Withey S, Montana G, editors. Longitudinal Detection of Radiological Abnormalities with Time-Modulated LSTM. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support; 2018 2018//; Cham: Springer International Publishing.
158. Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. (2398-6352 (Electronic)).
159. Sun Y, Wong AKC, Kamel MS. CLASSIFICATION OF IMBALANCED DATA: A REVIEW. International Journal of Pattern Recognition and Artificial Intelligence. 2009;23(04):687-719.

160. He HaM, Y. Imbalanced Learning: Foundations, Algorithms, and Applications, 1ste ed. Wiley-IEEE Press. 2013.
161. Ferri C, Hernández-Orallo J, Modroiu R. An experimental comparison of performance measures for classification. Pattern Recognition Letters. 2009;30(1):27-38.
162. Handelman GS, Kok HK, Chandra RV, Razavi AH, Huang S, Brooks M, et al. Peering Into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods. American Journal of Roentgenology. 2018;212(1):38-43.
163. Provost F, Fawcett T. Robust Classification for Imprecise Environments. Machine Learning. 2001;42(3):203-31.
164. Fatourechi M, Ward RK, Mason SG, Huggins J, Schlögl A, Birch GE, editors. Comparison of Evaluation Metrics in Classification Applications with Imbalanced Datasets. 2008 Seventh International Conference on Machine Learning and Applications; 2008 11-13 Dec. 2008.
165. Vickers A, J,, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. (1472-6947 (Electronic)).
166. Cook TS. The Importance of Imaging Informatics and Informaticists in the Implementation of AI. (1878-4046 (Electronic)).
167. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ. 2020;368:l6927.
168. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. (1558-254X (Electronic)).
169. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? Advances in Neural Information Processing Systems (NIPS). 2014;27.
170. Adaloglou N. Transfer learning in medical imaging: classification and segmentation [Internet]. The AI summer. Nov 2020. [Cited on April 26 2021]. Available from: https://theaisummer.com/medical-imaging-transfer-learning/.
171. Toll D, B,, Janssen KJ, Vergouwe Y, Moons K, G,. Validation, updating and impact of clinical prediction rules: a review. (1878-5921 (Electronic)).
172. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Quality & Safety. 2019;28(3):231.
173. Götte G, Öfele M, Wolters V. Beware of the Bias -How the Representation of Subgroups in Training Data Sets affects Binary Classifier Decisions2020.
174. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. (1538-3598 (Electronic)).
175. Rajpurkar PA-O, Irvin JA-O, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. (1549-1676 (Electronic)).
176. Williams D, Mohammed SA, Leavell J, Collins C. Race, socioeconomic status, and health: complexities, ongoing challenges, and research opportunities. (1749-6632 (Electronic)).
177. Bidgood WD, Jr., Horii SC, Prior FW, Van Syckle DE. Understanding and using DICOM, the data interchange standard for biomedical imaging. J Am Med Inform Assoc. 1997;4(3):199-212.
178. Aryanto KY, Oudkerk M, van Ooijen PM. Free DICOM de-identification tools in clinical research: functioning and safety of patient privacy. (1432-1084 (Electronic)).
179. Google-Healthcare. De-identifying DICOM data. https://cloud.google.com/healthcare/docs/how-tos/dicom-deidentify. Published 2019. Accessed April 2021.
180. Price WN, 2nd, Cohen IG. Privacy in the age of medical big data. (1546-170X (Electronic)).
181. Sheng VS PF, Ipeirotis PG. Get another label? improving data quality and data mining using multiple, noisy labelers. Proceedings of the 14th ACM SIGKDD international conference on

Knowledge discovery and data mining, Las Vegas, August 24–27, 2008. New York, NY: Association for Computing Machinery, 2008; 614–622.
182. European Society of R. ESR paper on structured reporting in radiology. Insights Imaging. 2018;9(1):1-7.
183. Kuhn, M. & Johnson, K. (2013). Applied predictive modeling. Springer.
184. Summers RM, Handwerker LR, Pickhardt PJ, Van Uitert RL, Deshpande KK, Yeshwant S, et al. Performance of a Previously Validated CT Colonography Computer-Aided Detection System in a New Patient Population. American Journal of Roentgenology. 2008;191(1):168-74.
185. Little MA, Varoquaux G, Saeb S, Lonini L, Jayaraman A, Mohr DC, et al. Using and understanding cross-validation strategies. Perspectives on Saeb et al. Gigascience. 2017;6(5):1-6.
186. Benner J. Cross-Validation and Hyperparameter Tuning: How to Optimise your Machine Learning Model [Internet]. Towards data science. Aug 2020. [Cited on April 22 2020] Available from: https://towardsdatascience.com/cross-validation-and-hyperparameter-tuning-how-to-optimise-your-machine-learning-model-13f005af9d7d [
187. Rahman MaD, DN., "Addressing the Class Imbalance Problem in Medical Datasets," International Journal of Machine Learning and Computing vol. 3, no. 2, pp. 224-228, 2013.
188. Zhao Y, Wong ZS-Y, Tsui KL. A Framework of Rebalancing Imbalanced Healthcare Data for Rare Events' Classification: A Case of Look-Alike Sound-Alike Mix-Up Incident Detection. Journal of Healthcare Engineering. 2018;2018:6275435.
189. Chawla N, Bowyer K, Hall L, Kegelmeyer WPJJAIR. SMOTE: Synthetic Minority Over-sampling Technique. 2002;16:321-57.
190. Thai-Nghe N, Gantner Z, Schmidt-Thieme L, editors. Cost-sensitive learning methods for imbalanced data. The 2010 International Joint Conference on Neural Networks (IJCNN); 2010 18-23 July 2010.

8. Appendices

Appendix 1: Search strategy per database

PubMed

("Coronary Angiography"[Majr] OR "Coronary Angiography"[ti] OR "Coronary Angiographies"[ti] OR "coronary angiogram*"[ti]) AND ("Artificial Intelligence"[Majr] OR "Artificial Intelligence"[ti] OR "AI"[tiab] OR "Image Processing, Computer-Assisted"[Majr] OR "automatic image analysis"[ti] OR "machine learning"[ti] OR "deep learning"[ti]) NOT ("CT"[ti] OR "compute*"[ti] OR "tomogr"[ti] OR "SPECT"[ti] OR "MR"[ti] OR "MRI"[ti] OR "magnetic"[ti])

Web of Science

TS=("Coronary Angiography" OR "Coronary Angiographies" OR "coronary angiogram*") AND TI=("Artificial Intelligence" OR "AI" OR "computer assisted Image Processing" OR "automatic image analysis" OR "machine learning" OR "deep learning") NOT TI=("CT" OR "compute*" OR "tomogr" OR "SPECT" OR "MR" OR "MRI" OR "magnetic")

IEEE

"All Metadata":coronary angiography AND ("All Metadata":artificial intelligence OR "All Metadata":deep learning OR "All Metadata":machine learning) NOT ("Document Title":"CT" OR "Document Title":"Computed tomography*")

EMBASE

('artificial intelligence':ti,ab,kw OR 'artificial intelligence'/exp OR 'deep learning':ti,ab,kw OR 'deep learning'/exp OR 'machine learning':ti,ab,kw OR 'machine learning'/exp) AND ('coronary angiography'/exp OR 'coronary angiography':ti,ab,kw) NOT ('computer assisted tomography'/exp OR 'nuclear magnetic resonance imaging'/exp)

Google Scolar

"coronary angiography*" "artificial intelligence"|"deep learning" -"computed tomography" -"SPECT" -"MRI" -"kidney"

Appendix 2: Interview questions

1. Have you ever heard of AI/ML/DL
    a. Yes
    b. No
2. How would you rate your knowledge on AI / ML /DL
    a. Never heard of it (1)
    b. Heard of it (2)
    c. Reasonable knowledge (3)
    d. Extensive knowledge (4)
    e. Expert (5)
3. Do you expect a role for AI in the future of interventional cardiology?
    a. Yes
    b. No
    c. Maybe
4. To what extend will AI influence interventional cardiology
    a. Within 5 years
        i. Not at all (1), barely (2), reasonably (3), extensively (4), very extensively (5)
    b. In 5 – 10 years
        i. Not at all (1), barely (2), reasonably (3), extensively (4), very extensively (5)
    c. In more than 10 years
        i. Not at all (1), barely (2), reasonably (3), extensively (4), very extensively (5)
5. Within how many years do you expect the first effects in clinical practice? (open)
    a. ..
    b. Never
6. What will be the effect of AI on interventional cardiology?
    a. Very negative (1)
    b. Negative (2)
    c. No effect (3)
    d. Positive (4)
    e. Very positive (5)
7. Why do you expect a positive effect of AI? Multiple answers possible (skip when answered 6 with a, b or c)
    a. Objective choice for intervention
    b. Improvement of strategy during PCI
    c. Improved diagnostics
    d. Optimization of workflow, logistics and planning outside the cathlab
    e. Reduction of costs
    f. Other: …
8. Why will AI not improve interventional cardiology Multiple answers possible (skip when answered 6 with d or e)
    a. Not applicable in clinical practice
    b. No additional value
    c. Ethical or legal problems
    d. Safety issues
    e. Costs
    f. Other: ..

9. Are you planning on learning more about this subject?
    a. Yes
    b. No
10. Would you be prepared to use AI in clinical practice? (in a safe way)
    a. Yes
    b. No
11. Would you be interested in working with computer or data scientists to develop an AI model?'
    a. Yes
    b. No
12. Which 4 challenges for implementation of AI in clinical practice are most relevant?
    a. High costs for software development
    b. High costs for software purchase
    c. Lack of confidence in AI by medical staff
    d. Lack of knowledge on AI by medical staff
    e. Lack of high quality image data
    f. Lack of high quality image labels
    g. Lack of generalizability
    h. Safety issues
    i. Ethical or legal problems
    j. Limited digital infrastructure in hospitals

Potential value of AI in acute setting

Subject	Diagnostics	Therapy
STEMI MINOCA		
NSTEMI MINOCA		
SCAD		
Lesion significance		
Stent choice		
Choice for intravascular imaging		
Complication risk		

Potential value of AI in elective setting

Subject	Diagnostics	Therapy
Coronary microvascular disease		
Vasospastic angina		
Myocardial bridge		
Generalised diffuse atherosclerosis		
Intermediary stenosis		
Lesion significance		
Stent choice		
Bifurcation stenting approach		
Choice for intravascular imaging		
Predicting restenosis		
Complication risk		
Management of iatrogenic dissections		

Appendix 3: Python scripts

3.1 DICOM to PNG

```
import os
from os import path
import cv2
import pydicom
import fnmatch
import numpy as np
from PIL import Image

# Directories
inputdir = r'K:\TM20-21\TM3_WvanderLoo\ICA_images'
outputdir = r'E:\WoutervanLoo\ICA_png'

# Check if output directory already exists
if path.exists(outputdir) == 0:
    os.mkdir(outputdir)  # create new directory for output
else:
    print('Directory already created')

os.chdir(inputdir)  # Change to directory with images
FileNamesDic = fnmatch.filter(os.listdir(inputdir), '*dic')
NumberOfFiles = range(len(FileNamesDic))


for i in NumberOfFiles:
    # create a folder per patient
    print(i)
    outputdir_seq1 = str(outputdir + '\\' + FileNamesDic[i])
    outputdir_seq = outputdir_seq1.replace(".dic", "")

# Check if patient directory already exists
    if path.exists(outputdir_seq) == 0:
        os.mkdir(outputdir_seq)  # create new dir  per patient
    else:
        print('Directory per patient already created')

    os.chdir(inputdir)
    ds = pydicom.read_file(FileNamesDic[i])
    img = ds.pixel_array
    img = img.astype('float64')
    img /= img.max()/255.0

    NumberOfFrames = range(np.shape(img)[0])

    os.chdir(outputdir_seq)
```

```
#   Convert all individual frames to png files
    for f in NumberOfFrames:
        Frame = img[f]
        FileNamesDic_Frame = FileNamesDic[i].replace(
            '.dic', '_'+ str(NumberOfFrames[f]) + '.dic')    # Create filena
mes PatientID_date_series_image_frame
        PNG_FileName = FileNamesDic_Frame.replace('.dic', '.png')
        cv2.imwrite(PNG_FileName, Frame)                     # save as png

        # Resize and convert to grayscale
        PNG_resized = Image.open(PNG_FileName)
        PNG_resized = PNG_resized.resize((512,512))
        PNG_resized_gray = PNG_resized.convert('L')
        PNG_resized_gray.save(PNG_FileName)
```

3.2 Frame selection

```
import os
from os import path
import pandas as pd
import shutil

inputFile = r'E:\WoutervanLoo\Queries\frame_selection.xls'
PNGdir = r'E:\WoutervanLoo\ICA_png'

# load .xls-file with data per case
# create series with info
df = pd.read_excel(inputFile, dtype='object')
PatientNumber = df['Patientnumber']
CathDate = df['Cath date']
Seq = df['number of sequence with best view']
ImageNumber = df['Image number']
QCA_frame = df['Python frame']
Label = df['Significance']

# Create a list with all the patient directories
patientdir = []
for x in range(0,len(df)):
    patientdir1 =  str(PNGdir + '\\' + str(PatientNumber[x]) + '_' + str(Ca
thDate[x])
                   + '_' + str(Seq[x]) + '_' + str(ImageNumber[x]))
    patientdir.append(patientdir1)
    patientdir[x] = patientdir[x].replace('-', '')
    patientdir[x] = patientdir[x].replace(' 00:00:00', '')

# create a list with all the filenames for the QCA frame
QCA_frame_filename = []
for x in range(0,len(df)):
    QCA_frame_filename1 = str(str(PatientNumber[x]) + '_' + str(CathDate[x]
)
```

```
                + '_' + str(Seq[x]) + '_' + str(ImageNumber[x]) + '_' + st
r(QCA_frame[x]) + '.png')
    QCA_frame_filename.append(QCA_frame_filename1)
    QCA_frame_filename[x] = QCA_frame_filename[x].replace('-', '')
    QCA_frame_filename[x] = QCA_frame_filename[x].replace(' 00:00:00', '')

# Move the selected files to the designated directory
QCA_outputdir = r'E:\WoutervanLoo\ICA_png\QCA_frames'

if path.exists(QCA_outputdir):
    print('QCA output directory already exists')
else:
    os.mkdir(QCA_outputdir)

# Check if direcotry for specific patients exists
for x in range(0,len(df)):
    if path.exists(patientdir[x]):
        os.chdir(patientdir[x])
    else:
        print(patientdir[x])
    if path.exists(QCA_frame_filename[x]):
        shutil.copy(QCA_frame_filename[x], QCA_outputdir)       # copy fram
e to QCA_frame folder
    else:
        print(QCA_frame_filename[x])

# names for directories per lavel
QCA_frames_dir = r'E:\WoutervanLoo\ICA_png\QCA_frames'
QCA_frames_sign = str(QCA_frames_dir + "/" + 'significant')
QCA_frames_insign = str(QCA_frames_dir + "/" + 'insignificant')

if path.exists(QCA_frames_sign):
    print('QCA_frames_sign dir already created')
else:
    os.mkdir(QCA_frames_sign)

if path.exists(QCA_frames_insign):
    print('QCA_frames_insign dir already created')
else:
    os.mkdir(QCA_frames_insign)

# get indices for label
Idx1 = [i for i, x in enumerate(Label) if x ==1]    # significant CAD
Idx0 = [i for i, x in enumerate(Label) if x ==0]    # insignificant CAD

os.chdir(QCA_frames_dir)
# Check if file exists and copy to designated directory
for i in range(0,len(Idx1)):
   if path.exists(QCA_frame_filename[Idx1[i]]):
```

```
        shutil.copy(QCA_frame_filename[Idx1[i]], QCA_frames_sign)
    else:
         print(QCA_frame_filename[Idx1[i]])

for i in range(0, len(Idx0)):
    if path.exists(QCA_frame_filename[Idx0[i]]):
        shutil.copy(QCA_frame_filename[Idx0[i]], QCA_frames_insign)
    else:
        print(QCA_frame_filename[Idx0[i]])
```

3.3 Data split

```
import numpy as np
from sklearn.model_selection import train_test_split
import fnmatch
import os
from os import path
import shutil

# define locations of files
inputdir_sign = r'E:\WoutervanLoo\ICA_png\significant'
inputdir_insign = r'E:\WoutervanLoo\ICA_png\insignificant'

# names for directories
outputdir_sign_train = r'E:\WoutervanLoo\training\sign'
outputdir_insign_train = r'E:\WoutervanLoo\training\insign'
outputdir_sign_test = r'E:\WoutervanLoo\testing\sign'
outputdir_insign_test = r'E:\WoutervanLoo\testing\insign'

# find all .png files in folders
FileNames_sign = fnmatch.filter(os.listdir(inputdir_sign), '*png')
FileNames_insign = fnmatch.filter(os.listdir(inputdir_insign), '*png')

# Create directories for training and testing
os.chdir(r'E:\WoutervanLoo')
if path.exists(outputdir_sign_train) ==0:
    os.makedirs(outputdir_sign_train)
else:
    print('training sign alreasy exists')
if path.exists(outputdir_insign_train) ==0:
    os.makedirs(outputdir_insign_train)
else:
    print('training insign alreasy exists')
if path.exists(outputdir_sign_test) ==0:
    os.makedirs(outputdir_sign_test)
else:
    print('Testing sign alreasy exists')
if path.exists(outputdir_insign_test) ==0:
    os.makedirs(outputdir_insign_test)
else:
```

```
    print('Testing insign alreasy exists')

# randomly assign cases to testing and training while maintain balance
sign_train, sign_test = train_test_split(FileNames_sign, test_size=0.2
random_state=420)

insign_train, insign_test = train_test_split(FileNames_insign, test_size=0.
2, random_state=420)

# copy files to intended dirs
os.chdir(inputdir_sign)

for i in range(0,len(sign_train)):
   if path.exists(sign_train[i]):
       shutil.copy(sign_train[i], outputdir_sign_train)
   else:
        print('Missing file')

for i in range(0,len(sign_test)):
   if path.exists(sign_test[i]):
       shutil.copy(sign_test[i], outputdir_sign_test)
   else:
        print('Missing file')

os.chdir(inputdir_insign)

for i in range(0,len(insign_train)):
   if path.exists(insign_train[i]):
       shutil.copy(insign_train[i], outputdir_insign_train)
   else:
        print('Missing file')

for i in range(0,len(insign_test)):
   if path.exists(insign_test[i]):
       shutil.copy(insign_test[i], outputdir_insign_test)
   else:
        print('Missing file')
```

3.4 Rebalance

```
# Rebalance data to 50/50 by minority oversampling
import imblearn
from imblearn.over_sampling import RandomOverSampler
import os
import pandas as pd
import numpy as np
import random
import shutil
from PIL import Image

outputdir = r'E:\WoutervanLoo\Kfold_manual_VGG3\kf5\train\sign'
oversample = RandomOverSampler(sampling_strategy='minority')

#Load filenames from existing directories
FileNames_sign = os.listdir(outputdir)
FileNames_insign = os.listdir(r'E:\WoutervanLoo\Kfold_manual_VGG3\kf5\train
\insign')

N_sign = len(FileNames_sign)
N_insign = len(FileNames_insign)
Oversampling_sign_N = N_insign-N_sign

FileNames_sign_b = random.sample(FileNames_sign, Oversampling_sign_N)

os.chdir(outputdir)
for i in range(0, Oversampling_sign_N):
    FileName_oversampled = FileNames_sign_b[i].replace('.png', '_2.png')
    shutil.copyfile(FileNames_sign_b[i], str(outputdir + '\\' + FileName_ov
ersampled))
```

3.5 Neworks

```
import keras,os
from keras.models import Sequential
from keras.layers import Dense, Conv2D, MaxPool2D, Flatten, BatchNormalizat
ion, Dropout
from keras.preprocessing.image import ImageDataGenerator
import numpy as np
from keras.optimizers import Adam
from keras.callbacks import ModelCheckpoint, EarlyStopping

# Loading data from specific directories
# Labeling images based on location and pass into network
BS = 5
TS = (512, 512)
FileName_model = "Network_name"
trdata = ImageDataGenerator(brightness_range=[0.5, 1.5], rotation_range=20,
```

```
                            width_shift_range =[-
5, 5], height_shift_range = [-5, 5],
                            shear_range=2.0, rescale=1/255.0 )
traindata = trdata.flow_from_directory(directory=r'E:\WoutervanLoo\Balance\
training',target_size=TS, color_mode="grayscale", batch_size=BS, class_mode
 ='categorical')

tsdata = ImageDataGenerator(rescale=1/255.0)
testdata = tsdata.flow_from_directory(directory= r'E:\WoutervanLoo\Balance\
testing',target_size=TS, color_mode="grayscale", batch_size=BS,class_mode =
'categorical')
SPE_T = np.ceil(len(traindata.filepaths)/BS)
VS = np.ceil(len(testdata.filepaths)/BS)
```

Network 1

```
model = Sequential()
# block 1
model.add(Conv2D(32, (3, 3), activation='relu', kernel_initializer='random_
normal', padding='same', input_shape=(512, 512, 1)))
model.add(BatchNormalization())

model.add(MaxPool2D((2, 2)))
model.add(BatchNormalization())

# block 2
model.add(Conv2D(64, (3, 3), activation='relu', kernel_initializer='random_
normal', padding='same'))
model.add(BatchNormalization())

model.add(MaxPool2D((2, 2)))
model.add(BatchNormalization())

# block 3
model.add(Conv2D(128, (3, 3), activation='relu', kernel_initializer='random
_normal', padding='same'))
model.add(BatchNormalization())

model.add(MaxPool2D((2, 2)))
model.add(BatchNormalization())

model.add(Flatten())
model.add(Dense(128, activation='relu', kernel_initializer='random_normal')
)
model.add(BatchNormalization())
model.add(Dropout(0.2))
model.add(Dense(2, activation='softmax'))
```

Network 2

```
# initialize VGG16 model
model = Sequential()

# 2x Convolution layer, BN, MP, BN
model.add(Conv2D(input_shape=(512,512,1),filters=64,kernel_size=(3,3), padd
ing="same", activation="relu"))
model.add(Conv2D(filters=64,kernel_size=(3,3),padding="same", activation="r
elu"))
model.add(BatchNormalization())
model.add(MaxPool2D(pool_size=(2,2),strides=(2,2)))
model.add(BatchNormalization())

# 2x Convolution layer, BN, MP, BN
model.add(Conv2D(filters=128, kernel_size=(3,3), padding="same", activation
="relu"))
model.add(Conv2D(filters=128, kernel_size=(3,3), padding="same", activation
="relu"))
model.add(BatchNormalization())
model.add(MaxPool2D(pool_size=(2,2),strides=(2,2)))
model.add(BatchNormalization())

# 2x Convolution layer, BN, MP, BN
model.add(Conv2D(filters=256, kernel_size=(3,3), padding="same", activation
="relu"))
model.add(Conv2D(filters=256, kernel_size=(3,3), padding="same", activation
="relu"))
model.add(BatchNormalization())
model.add(MaxPool2D(pool_size=(2,2),strides=(2,2))) #
model.add(BatchNormalization())

# 2x Conv, BN, MP, BN, MP, BN
model.add(Conv2D(filters=512, kernel_size=(3,3), padding="same", activation
="relu"))
model.add(Conv2D(filters=512, kernel_size=(3,3), padding="same", activation
="relu"))
model.add(BatchNormalization())
model.add(MaxPool2D(pool_size=(2,2),strides=(2,2)))
model.add(BatchNormalization())
model.add(MaxPool2D(pool_size=(2,2),strides=(2,2)))
model.add(BatchNormalization())

model.add(Flatten()) # Flatten the vector generated by convolutions
model.add(Dense(units=1024,activation="relu"))  # Dense layer of 1024 units
 (fully connected layer)
model.add(BatchNormalization())
model.add(Dropout(0.2)) # dropout layer with 0.2 dropout rate
model.add(Dense(units=2, activation="softmax"))
#%% Compile model
```

```
opt = Adam(lr=0.00001)    # Adam optimiser to reach global minima while tra
ining, learning rate 0.001
model.compile(optimizer=opt, loss=keras.losses.BinaryCrossentropy(from_logi
ts=True),metrics=['accuracy'])

model.summary()

checkpoint = ModelCheckpoint(FileName_model, monitor='val_accuracy', verbos
e=1,        # Save model and weights
                              save_best_only=True, save_weights_only=False,
                              mode='auto', period=1)
early = EarlyStopping(monitor='val_accuracy', min_delta=0, patience=40,
verbose=1, mode='auto') # Stop when plateau is reached
hist = model.fit_generator(steps_per_epoch=SPE_T, generator=traindata
                           validation_data= testdata, validation_steps=VS,e
pochs=100,callbacks=[checkpoint,early])
```

3.6 Prediction

```
import keras
import os
from keras.models import load_model
import matplotlib.pyplot as plt
import numpy as np
from PIL import Image
from sklearn.metrics import confusion_matrix
from keras.preprocessing.image import ImageDataGenerator

# load the trained model
os.chdir(r'C:\Users\coassistent')
model = load_model('VGG16.h5', compile=True) # load model

# Create iterator for data flow
# Set shuffle to false to get consistent results
BS = 1
TS = (512, 512)
tsdata = ImageDataGenerator(rescale=1/255.0)
testdata = tsdata.flow_from_directory(directory= r'E:\WoutervanLoo\Balance\
testing',target_size=TS, color_mode="grayscale",batch_size=BS,class_mode ='
categorical', shuffle=False)
model.summary()

# Predict classes on new input data
STEP_SIZE_TEST= testdata.n//testdata.batch_size
testdata.reset()
prediction = model.predict(testdata, steps=STEP_SIZE_TEST)

# Get class with highest probability
classes = np.argmax(prediction, axis=1)
```

```
cf_matrix = confusion_matrix(testdata.classes, classes) # create confusion
matrix
acc = (cf_matrix[0,0] + cf_matrix[1,1])/sum(cf_matrix.flatten())
```

3.8 Grad CAM visualization

```
# GRAD CAM implementation
from tensorflow.keras.models import Model
import numpy as np
import cv2
from vis.visualization import visualize_cam, overlay
import matplotlib.pyplot as plt
import matplotlib.cm as cm
from vis.utils import utils
from keras import activations
from keras.preprocessing.image import ImageDataGenerator
import os
from keras.models import load_model
import keras
from PIL import Image

trained_model = 'VGG16_BN_DO2.h5'
os.chdir(r'C:\Users\coassistent')
model = load_model(trained_model, compile=True)  # load trained model
model.summary() # look up name of final layer

layer_index = utils.find_layer_idx(model, 'dense_2') # Use final layer
model.layers[layer_index].activation = activations.linear
model = utils.apply_modifications(model)  # swap softmax layer for linear

# Get image for visalization
PNG_img = Image.open(r'E:/WoutervanLoo/testing/insign/FileName.png')

PNG_resized = PNG_img.resize((512,512)) # Reisze to desired image sizze
PNG_resized_gray = PNG_resized.convert('L')
img_np = np.asarray(PNG_resized_gray)
img_np = img_np*(1/255.0)
img_np = img_np.reshape(1,512, 512, 1) # stack images as batch

visualization = visualize_cam(model, layer_index, filter_indices=1, seed_in
put=img_np)

img = img_np.reshape(512,512) # reshape to normal image dimensiosn for visu
alization
alpha =0.2  #transparency of heatmap

heatmap = np.uint8(255*visualization)
jet = cm.get_cmap('jet')
jet_colors = jet(np.arange(256))[:, :3]
jet_heatmap = jet_colors[heatmap]
```

```
jet_heatmap = keras.preprocessing.image.array_to_img(jet_heatmap)
jet_heatmap = jet_heatmap.resize((img.shape[1], img.shape[0]))
jet_heatmap = keras.preprocessing.image.img_to_array(jet_heatmap)

# create rgb format to make images compatible
img_rgb = cv2.merge([img, img, img]) # Not needed when using 3-
channel images

# Superimpose the heatmap on original image
superimposed_img = jet_heatmap * alpha + (img_rgb*255)
superimposed_img = keras.preprocessing.image.array_to_img(superimposed_img)

# Create subplot
fig, (ax1, ax2, ax3) = plt.subplots(1,3)
fig.suptitle('GRAD CAM Visualization', y=0.80)
ax1.axis('off')
ax2.axis('off')
ax3.axis('off')
ax1.set_title('Original image')
ax1.imshow(img_rgb)
ax2.set_title('Heatmap')
ax2.imshow(jet_heatmap)
ax3.set_title('Overlay')
ax3.imshow(superimposed_img)
```

Appendix 4: Metrics and confusion matrices

$$\text{Accuracy} = \frac{\text{TP} + \text{FP}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{F1} - \text{score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

The layout of the confusion matrix from sklearn package differs from the conventional layout. Here the layout shown in table 8.4.1 is used. The insignificant CAD class is regarded as negative and the significant CAD class is regarded as positive.

Table 8.4.1 Layout of the confusion matrix

	Predicted negative	Predicted positive
Actual negative	True negatives	False positives
Actual positive	False negatives	True positives

Table 8.4.2 Confusion matrices for the six network configurations

N1	
28	4
8	10

N1_BN	
24	8
6	12

N1_BN_DO	
21	11
2	16

N2	
30	**2**
13	**5**

N2_BN	
21	11
6	12

N2_BN_DO	
23	**9**
13	**5**

Table 8.4.3 Confusion matrices for three channel images

N1	
21	7
8	14

N1_BN	
20	11
8	11

N1_BN_DO	
23	9
6	12