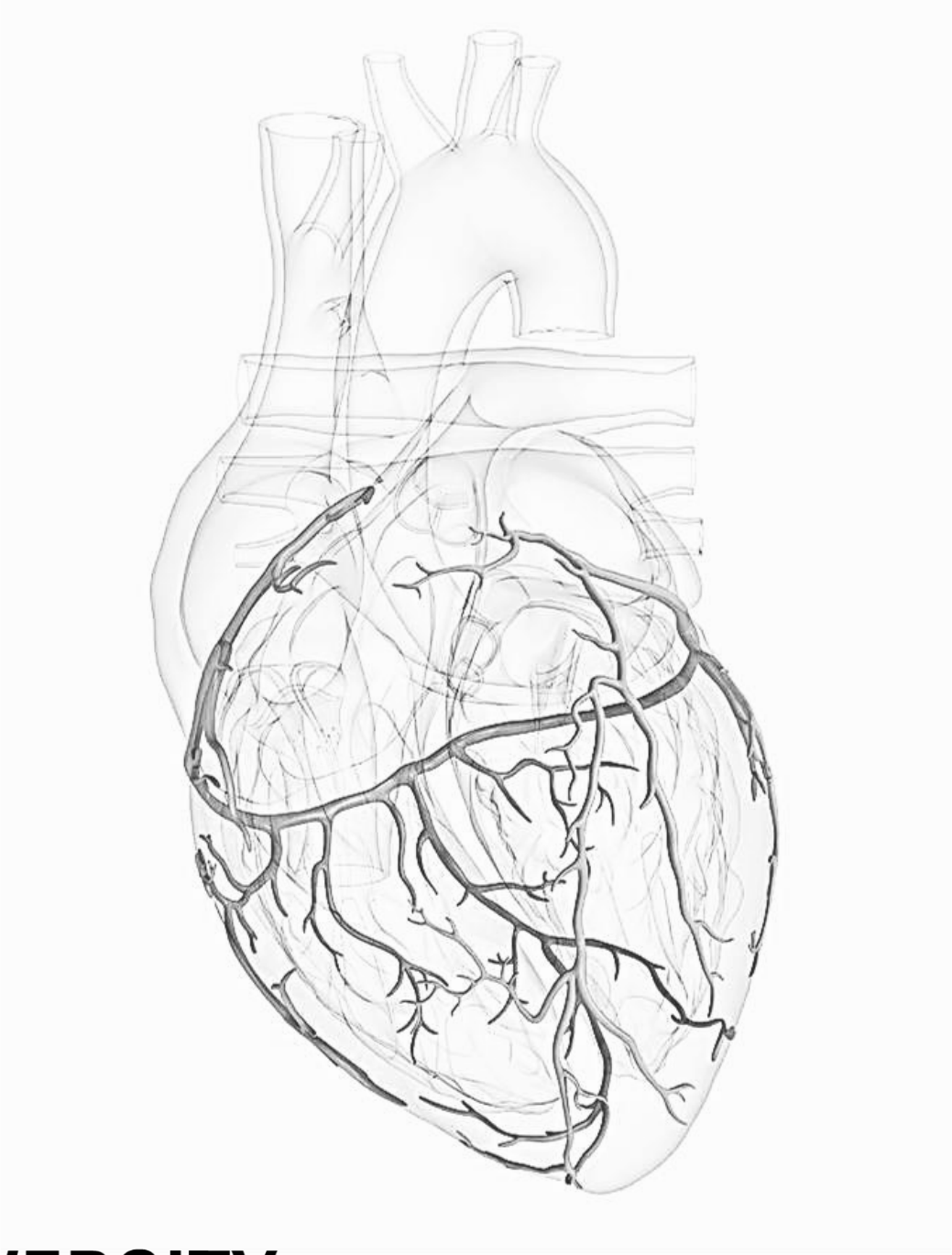
Applying machine learning to diagnose obstructive coronary artery disease

Jeroen Valk



UNIVERSITY OF TWENTE.

isala

UNIVERSITY OF TWENTE

A master's thesis for the degree of Technical Medicine

Applying machine learning to diagnose obstructive coronary artery disease

Author:
J.G. Valk

Graduation committee:
Prof. Dr. R.H.J.A Slart
Drs. B.N. Vendel, MD
Dr. C.O. Tan
Dr. J.D. van Dijk, MSc, MBA
Drs. N.S. Cramer Bornemann
Dr. J.M. Wolterink

01 Juli 2024

Abstract

Introduction. Risk stratification in patients with suspected chronic coronary syndromes is important to determine appropriate treatment. Decision making for expert readers is complex due to tens of variables to consider. Machine learning has demonstrated its suitability to achieve accurate diagnostic performance using a large number of features. The aim of this thesis is to develop and validate a machine learning-based model to diagnose obstructive coronary artery disease (oCAD) in patients without prior history of CAD using clinical data, medication, and imaging data from CACS and CCTA, and to compare the performance of this model to that of expert readers.

Method. We retrospectively included 1254 patients without a prior history of CAD who underwent CACS and CCTA. The entire dataset was split 4:1 into a training and test dataset, respectively. An Extra Trees Classifier (ETC) was developed on the training set using ten-fold stratified cross-validation and gridsearch for hyperparameter tuning. Boruta feature selection was utilized to minimize the impact of noise. The test dataset was used to compare the diagnostic performance of the model to the performance of expert readers. The primary endpoint was oCAD on invasive coronary angiographies (ICA).

Results. A total of 85 (6.8%) of the 1254 patients were diagnosed with oCAD. Imaging features were found to be most important to the predictive performance of the ETC model, with eight out of the nine selected features being derived from imaging. The ETC achieved a mean ROC-AUC of 0.94 on the training set and showed an ROC-AUC of 0.96 on the test set. The ETC model showed similar performance compared to the expert readers in accuracy (93% vs. 94%), precision (50% vs. 54%), and recall (94% vs 83%). McNemar Chi-Squared test indicated that there is no statistically significant difference between the performance of the model and that of the expert readers ($p > 0.99$).

Conclusion. We have developed and validated a machine learning model to diagnose oCAD. Our findings demonstrate that the ETC model performs comparable to expert readers in predicting oCAD. The model can assist expert readers in their decision-making by serving as a risk stratification tool. This study highlights the potential of machine learning in oCAD diagnosis.

Acknowledgements

Het afgelopen jaar was een prachtige ervaring waarin ik veel heb geleerd en heb gedaan. Dit alles zou niet mogelijk zijn geweest zonder de hulp en toewijding van een groot aantal mensen.

In het bijzonder wil ik mijn begeleider in het Isala bedanken, Brian, Joris en Jorn. Jullie waren altijd laagdrempelige vraagbakens, vriendelijk en enthousiast, en gaven mij de ruimte om de opdracht zelf in te vullen. Daarnaast wil ik jullie hartelijk danken voor de aanmoediging om een abstract te schrijven en de kans om naar een internationaal congres te gaan binnen mijn afstudeerstage. Niet te vergeten het aanzetten tot nadenken wat ik na mijn afstuderen zou willen doen.

Uiteraard wil ik ook Riemer en Can bedanken voor hun scherpe vragen en opmerkingen en het toezicht op de voortgang tijdens mijn afstuderen. Jelmer, ook jij verdient dank voor het aansluiten als mijn buitenlid en voor de aanmoediging om mijn horizon te verbreden op het gebied van modellen. Dankzij jullie heb ik mijn thesis naar een hoger niveau kunnen brengen.

Een speciale dank gaat uit naar Nicole, die mijn persoonlijke ontwikkeling de afgelopen twee jaar heeft meegemaakt en aangewakkerd. Je vele verdiepende vragen hebben me vaak aan het denken gezet en mijn persoonlijke ontwikkeling een extra boost gegeven.

Ik ben ook bijzonder dankbaar voor de hulp van Paul Hiemstra. Zonder jouw tijd en toewijding om me snel te helpen een model op te zetten, was het nooit gelukt om een abstract in te dienen voor ICNC-CT. Verder hebben onze vele discussies over verschillende aspecten van machine learning mijn interesse op dit gebied alleen maar versterkt.

Mijn dank gaat ook uit naar al mijn mede-(PhD) studenten in het Isala voor het overbruggen van mijn afstudeerstage, met de gezelligheid en het fungeren als mijn eerste vraagbaken. Zonder jullie was mijn tijd in Isala veel minder leuk en succesvol geweest.

Daarnaast wil ik mijn vriendin, Lilian, bedanken voor haar oneindige enthousiasme en aanmoedigingen. Ook wil ik Lilian en Lotte speciaal bedanken voor jullie kritische blik op spelling en grammatica.

Tot slot wil ik mijn ouders bedanken. Jullie onvoorwaardelijke steun is een drijfveer geweest gedurende mijn gehele studie. Ik zal altijd blijven streven om jullie trots te maken.

Contents

General introduction

Cardiovascular disease (CVD) is the leading cause of death worldwide, with coronary artery disease (CAD) being a major contributor [1,2]. In the Netherlands, CAD accounts for one in five deaths and was the fifth most common cause of death in 2021 [3]. CAD develops from atherosclerotic plaque buildup in coronary arteries, eventually causing narrowing and obstruction, leading to myocardial ischemia [4–6]. Symptoms include chest pain, fatigue, and shortness of breath [1,5,7].

Acute coronary syndrome (ACS) and chronic coronary syndrome (CCS) are CAD subgroups requiring different diagnostics [1,5]. ACS needs rapid assessment using electrocardiograms (ECGs), cardiac biomarkers, and invasive coronary angiography (ICA), while CCS involves risk stratification through non-invasive imaging. This research focuses on CCS diagnostics, specifically obstructive coronary artery disease (oCAD) diagnosis based on CT based calcium scoring (CACS) and coronary CT angiography (CCTA).

At Isala, a diagnostic workflow allows general practitioners to refer patients for CACS and CCTA based cardiac screening, to determine appropriate treatment. Decision making for expert readers is complex due to the amount of available intercorrelated data. Artificial intelligence and machine learning (ML) models can handle this amount of data, integrating imaging and clinical data for accurate diagnosis [8,9]. The aim of this thesis is to develop and validate a machine learning-based model to diagnose obstructive CAD in patients without prior history of CAD using clinical data, medication, and imaging data from CACS and CCTA, and to compare the performance of this model to that of expert readers.

The first chapter describes the clinical background of CAD and diagnostic pathway for oCAD diagnosis. The second chapter will serve as an introduction to machine learning and cover the three machine learning models that were used in this research. The third chapter, a standalone section designed as a setup for publication, covers the development and validation of the best performing machine learning model for predicting oCAD. The fourth and final chapter discusses future perspectives, clinical implementation, and recommendations for future research.

1. Clinical background

1.1 Pathophysiology of CAD

Coronary artery disease is characterized by the development of atherosclerotic plaque in the coronary arteries [10]. Atherosclerosis is a multifaceted pathological process characterized by the accumulation of lipids, inflammatory cells, and fibrous elements within arterial walls [11]. Atherosclerosis begins with endothelial dysfunction, allowing cholesterol infiltration and oxidation, triggering an inflammatory response and formation of fatty streaks [12]. Over time, smooth muscle cells migrate, proliferate, and form a fibrous cap over the fatty streak, initially preserving lumen size through vessel remodelling but eventually causing arterial narrowing, obstruction, and occlusion [4,5]. As the atherosclerotic plaque develops, it may undergo various changes, including calcification, lipid accumulation, and inflammation. Calcification involves the deposition of calcium salts within the plaque, rendering it harder and more brittle [13].

Vulnerable plaques, characterized by a thin fibrous cap, large lipid cores, and increased inflammation, are prone to rupture. Plaque rupture exposes thrombogenic material within the plaque to the bloodstream, causing platelet aggregation and thrombus formation. Resulting in an acutely reduced blood flow and oxygen supply to the myocardium, causing myocardial ischaemia [6]. Symptoms of acute myocardial ischemia include severe chest pain or discomfort that may radiate to the arm, jaw, neck, or back. The pain is often described as crushing, pressure-like, or squeezing and may be associated with sweating, nausea, vomiting, and dyspnoea [1]. Patients who suffer from a plaque rupture are treated according to the acute coronary syndrome guideline [1].

While stable plaques are less prone to rupture, they can still obstruct blood flow and cause symptoms of stable angina pectoris. Patients may experience chest discomfort or pain that is typically described as pressure, heaviness, or tightness in the chest. The pain is predictable and reproducible, occurring with similar intensity and duration upon exertion and relieved with rest or medication. Patients with these kinds of symptoms are treated according to the chronic coronary syndrome guidelines [5]. This research will focus on patients who suffer from CCS.

1.2 Chronic coronary syndromes guidelines and implementation in Isala

The diagnostic management of CCS according to the 2019 European Society of Cardiology (ESC) guidelines typically involves a step-by-step approach. Starting with clinical assessment which includes evaluating patient history, symptoms, and cardiovascular risk factors, followed by risk stratification using validated scoring systems like the Heart Score [14]. Followed by non-invasive testing such as exercise electrocardiography, stress echocardiography, CT scans, or nuclear imaging is then performed for ischemia detection and risk stratification based on patient characteristics and test availability. Finally, ICA is indicated for patients with high-risk features, positive non-invasive tests, or persistent symptoms despite optimal medical therapy, to confirm the diagnosis.

In the 2019 ESC CCS guidelines, CT calcium scoring, and coronary CT angiography play essential roles in risk stratification and assessment of coronary anatomy [5]. The recommendations of the ESC CCS guidelines from 2019 are well implemented in the clinical care in Isala. Special workflows have been developed to enhance the diagnostic accuracy of CAD. One of these workflows is focused on the diagnosis of oCAD for patients with CCS referred by general practitioners. The workflow is displayed in figure 1**.** Risk stratification is done based on age, CACS and CCTA. In the following paragraphs the basics of CACS and CCTA will be explained.

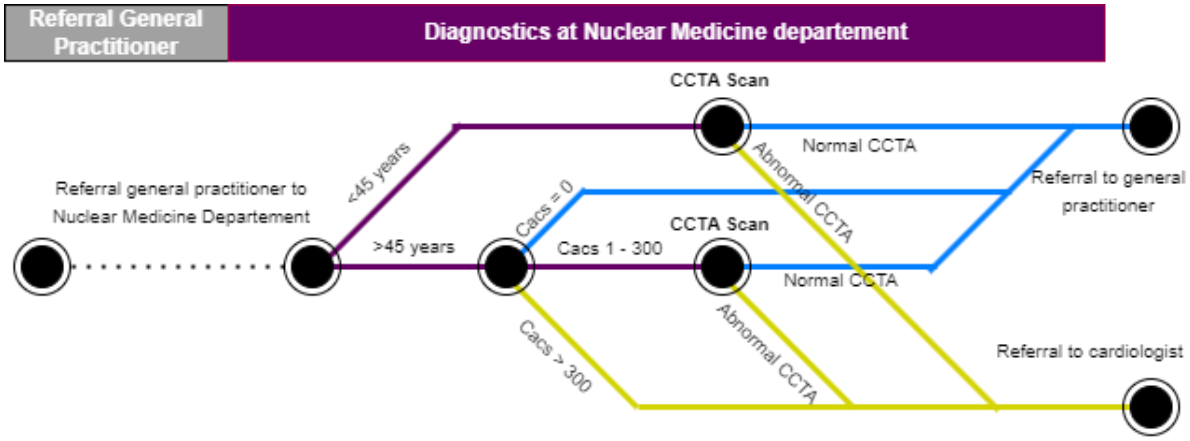


Figure 1 Diagnostic pathway for referral of patients with suspected CAD by general practitioners to the nuclear medicine department of Isala.

1.3 Coronary artery calcification scoring

Coronary artery calcification scoring, derived from an unenhanced low-dose CT scan, is vital for risk stratification in patients with suspected CCS, especially those lacking relevant medical history [15]. The severity of atherosclerosis is measured with CACS by calculating a weighted sum of lesions containing calcification, expressed in Agatston units (AU) [16]. The score's weighting depends on the density factor derived from the highest plaque attenuation in Hounsfield units. CACS is valuable due to its high negative predictive value, meaning a zero-score indicates a very low risk of oCAD [17]. This is why patients with a CACS score of 0 are referred back to their general practitioner, as can be seen in figure 1. However, scores between 1 and 300 AU suggest further diagnostic tests to differentiate between oCAD and non-obstructive CAD. Due to the general absence of calcification in plaques in patients below 45, CACS does not have predictive value in younger patients [18].

1.4 Coronary CT angiography

In contrast to CACS, Coronary CT angiography is the diagnostic modality of choice to evaluate the coronary anatomy [19]. Prior to the scan, iodine-based contrast medium is administered intravenously, enhancing contrast between the coronary arteries and adjacent structures. The resultant CTA data produces a three-dimensional volume of the coronaries, enabling comprehensive anatomical assessment. Through advanced post-processing techniques, detailed examination of coronaries can be achieved [20].

Heart rate management is crucial for CCTA scans to ensure high-quality images. Optimal heart rate, typically below 60 beats per minute, reduces motion artifacts and enhances image clarity [21]. Beta-blockers are commonly administered before the scan to achieve heart rate reduction. Coronary calcium can cause overestimation of stenosis due to blooming artifacts. These artifacts, resulting from the high density of calcified plaques, exaggerate the apparent size of the lesions and can obscure the true lumen of the artery. This overestimation is the reason for exclusion of CACS > 300 scores for CCTA imaging within the Isala, as seen in figure 1.

The Coronary Artery Disease - Reporting and Data System (CAD-RADS) is a standardized reporting system to classify the degree of stenosis, helping to guide subsequent patient management [22]. In figure 2 the CAD-RADS categories are displayed with the maximum degree of stenosis found during CCTA and advice for further diagnostic investigations. CAD-RADS scores of 3 and 4a particularly draw attention due to their inconclusive nature. For both these categories, the certainty of oCAD remains questionable. As such, patients with CAD-RADS 3 and 4a necessitate further diagnostic evaluation to delineate the exact extent of obstruction and to determine the most appropriate therapeutic course [22].

Category	Degree of Maximal Coronary Stenosis	Further Cardiac Investigation
CAD-RADS 0	0% (No plaque or stenosis)	None
CAD-RADS 1	1–24% (Minimal stenosis or plaque with no stenosis[b])	None
CAD-RADS 2	25–49% (Mild stenosis)	None
CAD-RADS 3	50–69% (Moderate stenosis)	Consider functional assessment[c]
CAD-RADS 4	**A** - 70–99% stenosis or **B** - Left main ≥50% or 3-vessel obstructive (≥70%) disease	A: Consider ICA[e] or functional assessment B: ICA is recommended
CAD-RADS 5	100% (total occlusion)	Consider ICA, functional and/or viability assessment
CAD-RADS N	Non-diagnostic study	Additional/alternative evaluation may be needed

Figure 2 CAD-RADS Reporting and data system for patients presenting with stable chest pain [22]

1.5 Gold standard evaluation of oCAD

Although CACS and CCTA are highly suitable for risk stratification, they are not capable of definitively confirming the presence of oCAD. Fractional Flow Reserve (FFR) guided invasive coronary angiography is the gold standard for the diagnostic evaluation of oCAD [23,24]. ICA offers direct visualization of the coronary vasculature, allowing for the precise delineation of luminal irregularities and stenosis. However, anatomical severity does not always correlate with functional significance. Consequently, FFR measured during ICA, assesses the hemodynamic relevance of stenosis by quantifying the pressure drop across it, thereby providing insights into its physiological impact [25]. An FFR value of ≤0.80 is considered indicative of a functionally significant stenosis [26].

2. Technical background

In recent years, machine learning has transformed the landscape of data analysis, enabling researchers and practitioners to extract valuable insights and make accurate predictions from complex datasets. This section will introduce machine learning models and methodology to optimize their predictive performance. Before delving into specific models, it is important to understand the fundamental principles upon which the methods that will be covered later are based, decision trees. Decision trees have an intuitive approach which mimics the human decision-making process, whereby decisions are recursively made by evaluating a series of binary choices based on feature values [27].

At each node of a decision tree, the algorithm selects the optimal feature and corresponding split point that maximizes the purity of the resulting subsets, typically measured by metrics such as Gini impurity or information gain [28]. This process continues until a stopping criterion is met, such as reaching a maximum tree depth or achieving minimum node purity. The resulting model is an advanced flowchart, which can be used to predict the classification of a new, unseen input.

Decision trees provide transparent insights into the decision-making process, allowing users to understand the underlying patterns and relationships within the data. Additionally, decision trees can handle both numerical and categorical features without requiring extensive preprocessing, making them well-suited for tabular data. However, decision trees are prone to overfitting, particularly on noisy datasets or datasets with many features, where they may exhibit poor generalization performance [29]. Decision trees can serve as the basis for more advanced algorithms which harness the collective predictive power of multiple trees to improve performance and generalizability.

2.1 Advanced machine learning models

The dataset used to train and validate the machine learning model in this study comprises data from 1,254 retrospectively included patients. The data is tabular and includes a mix of personal data (age, gender, BMI), clinical data (medication usage, Creatinine values), risk factors (family history, diabetes, etc), and imaging features from CACS and CCTA. This results in a comprehensive, highly intercorrelated, dataset with 33 features for each patient, containing dichotomous, categorical, and continuous variables. Selecting the appropriate machine learning model tailored for the dataset is crucial as it directly impacts the accuracy and reliability of the predictions [30]. A model that aligns well with the characteristics of the dataset maximizes its ability to extract patterns and relationships. This section introduces three models that are capable of handling this type of data.

2.1.1 Chosen Classifiers

Extreme Gradient Boosting (XGBoost) uses decision trees as its primary classifiers [31]. XGBoost repetitively improves the performance of multiple decision trees by focusing on instances where they misclassify data points, this process is called gradient boosting. It also involves training additional trees to address the differences between the actual and predicted outcomes, identified by the previous trees. Evaluation of the training set and prevention of overfitting are combined in the objective function. By minimizing the objective function, XGBoost optimizes the parameters of these decision trees. XGBoost is specifically designed for structured or tabular data, which makes it highly suited for the data at hand [32–34]. XGBoost is also designed to handle highly intercorrelated data, by selecting the most informative features for each split based on information gain, reducing the impact of noise. Furthermore, techniques to enhance generalizability are incorporated in XGBoost, such as tree pruning and shrinkage. These are essential in dataset with many features like our dataset. Finally, XGBoost is praised for its learning efficiency, due to parallel processing capabilities, reducing optimization and prediction times on large datasets.

Random Forest Classifiers (RFC) are ensemble learning techniques used for handling structured or tabular data with numerous features [35–38]. An RFC model operates by constructing numerous decision trees during training. Each tree is build independently using a random selection of features and data samples, the diversity among trees is increased by utilizing bootstrapping. The randomness that is build into the construction of the decision trees ensures that the impact of feature intercorrelation is limited and diverse patterns are recognised. The classification is predicted using a majority vote of all decision trees, reducing the impact of biases of individual trees. Random Forest Classifiers are less prone to overfitting as a result of the majority vote, thereby enhancing their generalizability. This approach reduces the influence of any single feature on the classification process, making them well-suited for handling intercorrelated features often found in tabular data [38]. Furthermore, Random Forest Classifiers excel in handling noisy or missing data, as well as categorical variables, without requiring extensive preprocessing. These characteristics ensure that an RFC model is well-suited for our dataset.

The Extra Trees Classifier (ETC), also known as Extremely Randomized Trees, is an extension of the Random Forest Classifier that introduces additional randomness during tree construction [39,40]. Unlike Random Forest, which selects the best split at each node based on a random subset of features, Extra Trees randomly selects splitting thresholds for each feature, without evaluating their optimality. This elevated level of randomness enhances diversity among the trees in the ensemble, making Extra Trees Classifier even less prone to overfitting. Additionally, like RFCs, ETC models provide insights into feature importance, helping to identify predictive features even when features are highly intercorrelated. Extra Trees Classifiers stand out in handling noisy or redundant features, as well as datasets with imbalanced class distributions, by averaging predictions from multiple trees to avoid bias and variance [41]. Their training process is less resource-intensive compared to other ensemble methods since they require less computational effort to find the optimal splitting thresholds.

2.1.2 Method of optimizing the model

To be able to achieve the best predictive performance of the chosen models it is imperative that the models are optimized for the inputted data. Proper optimization prevents overfitting and ensures that the models can reliably make predictions on unseen data. The dataset is pre-processed using Pandas to ensure the data is clean and ready for analysis. In the following paragraph a method is described that is used to optimize the chosen models. The resulting performance evaluation of the three models can be found in appendix A.

The first step in model development is to divide the dataset into training and test sets to assess model performance. A stratified split method ensures that class proportions are maintained in both sets, especially crucial for imbalanced datasets, like the dataset at hand. This approach prevents a bias in model evaluation and ensures representative samples for training and testing. The dataset is split in a 4:1 ratio between the training and test set [42]. Since the training set is split again, for cross validation, the test set is referred to as the validation set.

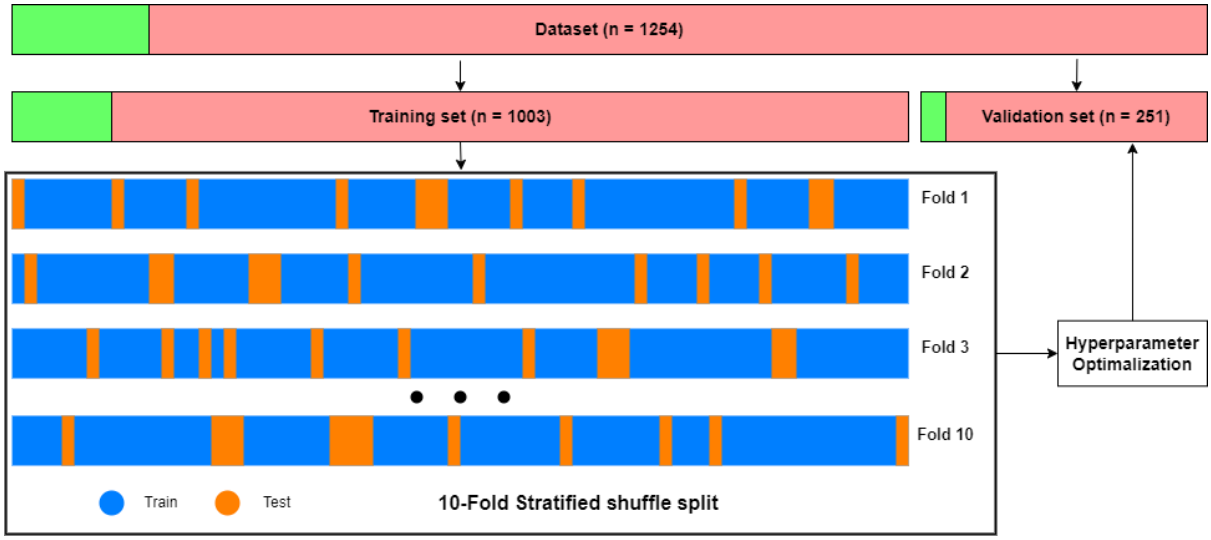


Figure 3 Visualization of the cross-validation procedure used to develop the models. The entire dataset is initially split in a 4:1 ratio between the training and validation set. The training set it used to optimize hyperparameters using cross-validation. The validation set is used to evaluate the performance in an unseen population.

To optimize model hyperparameters and avoid overfitting, a 10-fold cross-validation strategy is employed in combination with GridsearchCV (sci-kit learn, v1.5.0), which searches a specified hyperparameter grid to find the best model configuration [43,44]. The stratified shuffle split creates the opportunity to split the training set into infinite folds, while maintaining the wished proportion of test set, compared to the x-fold stratified split. The proportion of the test set is set to 20% each split. Figure 3 provides a complete overview of the splitting process.

A combination of four performance metrics is used to gain a broad understanding of the performance of the model. Accuracy measures the proportion of correctly classified instances, while precision quantifies the proportion of true positives among all predicted positives. Recall, also known as sensitivity, measures the proportion of true positives correctly identified by the model. Receiver operating characteristic area under the curve (ROC-AUC) evaluates the model's classification ability across different threshold values, providing a comprehensive assessment of the classifier's performance. The combination of these four metrics gains insight about probability, positive and negative predictive performance. Hyperparameters are chosen that scored best on the combination of these four metrics.

Finally, feature selection is crucial for identifying relevant features and reducing model complexity, by reducing noise. Boruta is a feature selection algorithm that evaluates the importance of each feature by comparing it with randomly generated shadow features [45]. Features are considered important if their importance scores significantly exceed those of shadow features. Features that are ranked less important than the highest scoring shadow feature are discarded. The Boruta algorithm are executed on five folds of the training set, to reduce bias of a singular fold. Features that are ranked important in three or more folds are included in the final selection.

3. Applying machine learning to diagnose obstructive coronary artery disease

3.1 Introduction

Worldwide cardiovascular disease (CVD) is the most common cause of death [1]. In the Netherlands one of every five deaths is related to CVD, with coronary artery disease (CAD) being the largest contributor to CVD mortality [2]. CAD is the fifth most common cause of death in the Netherlands in 2021. CAD is characterized by the development of atherosclerotic plaque in the coronary arteries [3,4]. In early stages of plaque development, lumen size tends to be preserved by vessel remodelling. Eventually, the continuous build-up of plaque within the intima of the vessel causes narrowing, obstruction, and occlusion in the arteries [5]. Resulting in a reduced blood flow and oxygen supply, causing myocardial ischemia [6]. Symptoms of myocardial ischemia include chest pain radiating to the jaw, shoulder blades and arms, fatigue, and shortness of breath [1,5,7].

Acute coronary syndrome (ACS) and chronic coronary syndrome (CCS) represent distinct subgroups within the spectrum of CAD, necessitating different diagnostic approaches [1,5]. In the context of suspected ACS, rapid assessment to determine the presence of myocardial ischemia or infarction is needed. Cardiac biomarkers and electrocardiograms (ECG) play a central role in the initial evaluation [1]. Patients with suspected CCS are primarily directed towards non-pharmacological interventions, emphasizing modifiable risk factor optimization. Concurrently, pharmacotherapy may be employed to not only mitigate risk factor intensification but also limit disease progression. For symptomatic patients with suspected CCS, the diagnostic focus shifts towards risk stratification using non-invasive cardiac imaging to evaluate the functional significance of coronary lesions [5].

Within the Isala, a diagnostic workflow is created for risk stratification of symptomatic CCS patients, without prior history of CAD. Risk stratification consists of CT based calcium scoring (CACS) and coronary CT angiography (CCTA) scans. The goal of the diagnostic workflow is to either refer a patient back to their general practitioner with minimal chance of obstructive CAD (oCAD) or refer them to a cardiologist for management of CAD, as can be seen in figure 4. Decision making for expert readers is complex due to tens of variables to consider, imaging data, risk factors, medication usage and clinical data, which are interrelated via complicated nonlinear relationships.

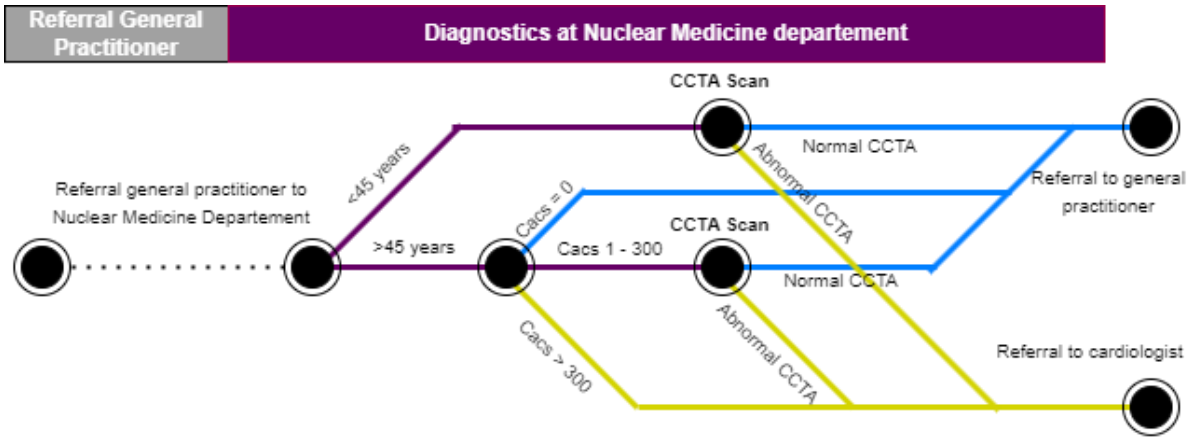


Figure 4 Diagnostic pathway for referral of patients with suspected CAD by general practitioners to the nuclear medicine department of Isala.

Artificial intelligence, in particular machine learning (ML) models, have been shown suitable to assess many features and are capable of modelling complex non-linear relations between these features, to finally result in an accurate diagnosis [8,9]. Previous studies have shown the potential of integrating imaging derived features with clinical data in ML-based risk prediction models, within the field of nuclear cardiology [46–48]. The aim of this thesis is to develop and validate a machine learning-based model to diagnose oCAD in patients without prior history of CAD using clinical data, medication, and imaging data from CACS and CCTA, and to compare the performance of this model to that of expert readers.

3.2 Method

3.2.1 Study design

We retrospectively included a cohort of 1254 patients, who were referred by general practitioners and underwent CACS and CCTA between January 1, 2018, and November 1, 2022. None of the patients had prior history of CAD, which was defined as previous myocardial infarction, percutaneous coronary intervention (PCI), or coronary artery bypass grafting (CABG). Patient history, characteristics, and clinical data were collected through a review of medical records. As this study was retrospective in nature, approval from the medical ethics committee was not required in accordance with Dutch law. Nonetheless, all patients provided written informed consent for utilization of their data for research purposes.

3.2.2. CACS and CCTA acquisition

CACS scans were performed using either a 64-slice CT scanner (Light-Speed VCT XT, GE Healthcare) or a 256-slice CT scanner (Revolution Apex, GE Healthcare). An unenhanced ECG-gated scan, triggered at 75% of the R-R interval, was obtained with the following parameters: 2.5 mm slice thickness, 120 kV tube voltage, and 125-250 mA tube current based on patient size. Dedicated software (SmartScore, GE Healthcare) calculated the CAC score per vessel using the Agatston criteria, with calcifications manually assigned to vessels by experienced operators.

Beta blockers were used for heart rate management in preparation for CCTA scan, maintaining rates under 60 for the 64-slice scanner and under 70 for the 256-slice scanner, with 0.4 mg nitroglycerin sublingual for vasodilation. CCTA scans were performed using either a 64-slice CT scanner (Light-Speed VCT XT, GE Healthcare) or a 256-slice CT scanner (Revolution Apex, GE Healthcare). Coronary arteries were prospectively scanned with ECG-triggering, using the kV-assist technique to automatically select tube voltage (80, 100, or 120 kV) based on patient size. Scanning parameters included a rotation time of 0.28s, collimation of 64 or 256 x 0.625 mm, a 512 x 512 reconstruction matrix, and detector coverage of 12 to 16 cm, employing bolus tracking. A low osmotic, non-ionic contrast medium (350 mg/ml Optiray or Iomeron) was administered based on patient size, followed by a 20 ml saline flush. The Advances Workstation (AW, GE Healthcare), with integrated Smart Phase for optimal phase selection, was used for image reconstruction and optimization.

3.2.3 Machine learning

Data preparation, model development and comparison to expert readers were implemented using Python (Python, v3.12, Python Software Foundation). The Extra Trees Classifier (ETC) from the scikit Learn library (v1.5.0) was utilized for the prediction of the presence of oCAD [46]. The dataset was randomly split into a training and test set with a 4:1 ratio, stratified by the occurrence of obstructive CAD. Each patient was characterized by an array of features, including CCTA measurements, CACS, and various other clinical features, shown in table 1**.** The list of features consists of dichotomous, categorical, and continuous variables. Normalization of the continuous variables was not performed, to prevent possible information loss.

The ETC was optimized on the training set, using hyperparameter optimization via gridsearch in combination with a 10-fold stratified shuffle split, as can be seen in figure 5. The performance of the classifier was evaluated using accuracy, precision, recall, and receiver operating characteristic area under the curve (ROC-AUC) [47]. For ROC-AUC analysis, a five-fold stratified split was introduced to show the performance of the model on different splits of the model. The hyperparameters were optimized for best average score of these performance metrics. Next, features were ranked for their importance. With Boruta feature selection, features were removed that were less important than noise, reducing noise and the complexity of the model. The Boruta algorithm was run on five stratified folds of the training set, to decrease the impact of the split and enhance the generalizability of the model after feature selection. Features selected by the Boruta algorithm in three or more folds were included in the final feature selection. The model with the highest average score of the performance metrics was evaluated on the test set.

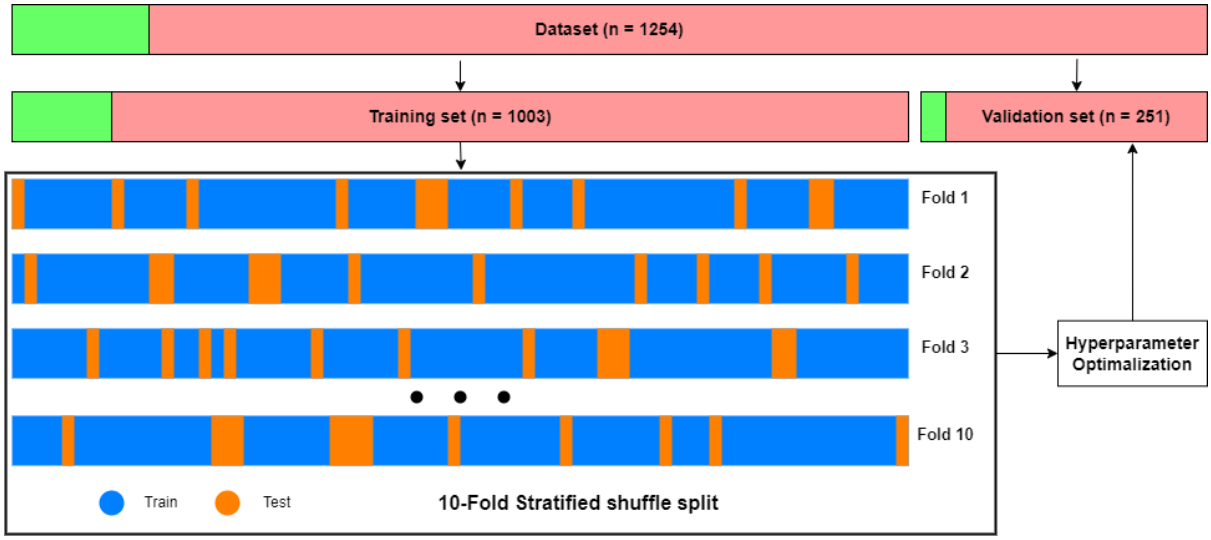


Figure 5 Visualization of the cross-validation procedure used to develop the models. The entire dataset is initially split in a 4:1 ratio between the training and validation set. The training set it used to optimize hyperparameters using cross-validation. The validation set is used to evaluate the performance in an unseen population.

The prediction of the model on the test set was compared to the performance of the expert readers. The expert readers, a cardiologist, and a nuclear medicine physician, assessed each patient after CACS and CCTA and categorized patients as having or not having oCAD. The predictive performance of the ETC was compared to the assessment of the expert readers. The performance of expert readers was assessed using a threshold of 120 days after CCTA examination. Diagnosis of oCAD after 120 days was noted as false negative prediction for the expert readers, due to late revascularization [48,49]. A reference standard was utilized to evaluate the diagnostic accuracy of an ML-based model in diagnosing oCAD. Patients were categorized as having oCAD if their follow-up included a conclusive ICA revealing oCAD, defined by a stenosis with a FFR ≤ 0.8 or underwent PCI or CABG procedures.

3.2.4 Statistical Analysis

Statistical analysis was performed using IBM SPSS (IBM SPSS Statistics for Windows, Version 28.0.1.0, Armonk, NY, IBM Corp). Differences in patient characteristics between the training and test set were evaluated with the Chi-squared test or Mann-Whitney U test. A statistical significance was set to 0.05. DeLong's test was used to compare ROC curves [50]. The performance of the ETC was compared to the assessment of the expert readers using McNemars Chi-squared test, the statistical significance was set to 0.05.

3.3 Results

3.3.1. Study population.

1254 patient were included from a large nuclear cardiology databases, the inclusion flowchart is displayed in figure 6. Of these patients, 163 underwent ICA interventions. 85 (6.8%) patients were diagnosed with obstructive CAD. 130 of the 163 ICAs were performed withing 120 days after the CCTA scan. 33 (20%) interventions were performed after 120 days, of which 13 patients were diagnosed with oCAD. 11 deaths were reported during follow-up, non with a confirmed cardiac cause. The median follow-up was 2.9 years (IQR 1.6 – 3.4 years). No significant differences were found between the characteristics of the patients in the training set compared to the patients in the test set, as can be seen in table 1.

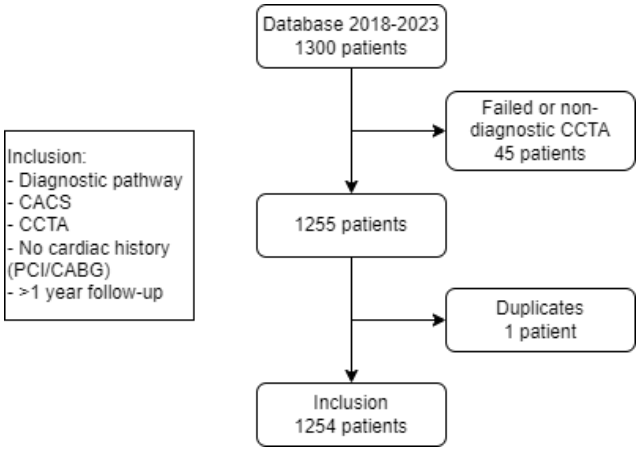


Figure 6 Inclusion flowchart. 45 patients were removed from the initial search due to failed or non diagnostic CCTA and one duplicate was removed from the second database.

3.3.2 Model development

The best performance of the ETC model was achieved using a set of eight hyperparameters. The hyperparameters were set as follows: class weights to balanced subsample, maximum depth to 10, maximum features to 'sqrt', minimum samples per split to 10, minimum samples per leaf to 1, number of estimators to 200, bootstrapping to true, and the criterion to Gini impurity. After hyperparameter optimization a mean-AUC of 0.92 was achieved using five-fold cross validation on the training set. The resulting ROC curves for the five-fold stratified split on the training set is displayed in figure 7. Features were ranked in order of importance on a single fold of the training set. The input of imaging features was most important to the predictive performance of the model. The top four features were acquired from CCTA. Most medication features, except aspirin prescription, were ranked as least predictive. The ranked feature importance of the model before feature selection can be seen in figure 8.

Table 1 Comparison of patient characteristics and all features between the training and validation *set. Continuous features are characterized by the mean values and standard deviations. Dichotomous and categorical features are characterized by the percentage of occurrence.*

Variable		Training set (1003 patients)	Test set (251 patients)	p-value
Male		54%	53%	0,809
Age		59.0 ± 10.16	59.5 ± 9.85	0,505
Weight		83.7 ± 16.28	84.2 ± 15.49	0,586
Height		173.8 ± 9.64	174.3 ± 10.20	0,421
BMI		27.2 ± 4.59	27.1 ± 4.09	0,994
Total Calciumscore		67.8 ± 95.13	73.2 ± 92.17	0,367
Calciumscore LM		4.6 ± 16.51	5.6 ± 18.68	0,335
Calciumscore LAD		40.5 ± 58.58	44.1 ± 54.45	0,346
Calciumscore CX		8.2 ± 22.50	8.2 ± 25.10	0,423
Calciumscore RCA		13.9 ± 33.82	15.1 ± 36.55	0,810
Degree of stenosis LM CTA				0,411
	0% stenosis	86%	87%	
	1 - 49% stenosis	13%	13%	
	50 - 69% stenosis	1%	0%	
	>70% stenosis	0%	0%	
Degree of stenosis LAD CTA				0,317
	0% stenosis	24%	24%	
	1 - 49% stenosis	65%	62%	
	50 - 69% stenosis	8%	11%	
	>70% stenosis	3%	2%	
Degree of stenosis CX CTA				0,482
	0% stenosis	70%	69%	
	1 - 49% stenosis	28%	28%	
	50 - 69% stenosis	2%	2%	
	>70% stenosis	0%	1%	
Degree of stenosis RCA CTA				0,551
	0% stenosis	61%	59%	
	1 - 49% stenosis	35%	37%	
	50 - 69% stenosis	2%	3%	
	>70% stenosis	1%	0%	
CAD-RADS				0,394
	CAD-RADS 0	11%	15%	
	CAD-RADS 1	49%	43%	
	CAD-RADS 2	22%	27%	
	CAD-RADS 3	9%	10%	
	CAD-RADS 4a	5%	5%	
	CAD-RADS 4b	1%	0%	
	CAD-RADS 5	0%	0%	
Creatinine		79.4 ± 16.65	79.7 ± 14.04	0,450
Clopidogrel		4%	5%	0,472
Aspirin		12%	12%	0,763
Acenocoumarol		1%	1%	0,868
DOACs		4%	4%	0,998
Beta-blocker		39%	36%	0,393
Ace inhibitor		11%	10%	0,620
Ca-channel blocker		13%	12%	0,639
Statin		19%	20%	0,808
Diuretics		2%	2%	0,772
Ezetimibe		1%	2%	0,261
Family history		48%	47%	0,789
Hypertension		44%	40%	0,261
Smoking		41%	38%	0,309
Hyperlipidaemia		38%	35%	0,359
Diabetes Mellitus		8%	9%	0,568
COPD		7%	8%	0,510
ICA		12%	10%	0,404
Obstructive CAD		7%	7%	0,997

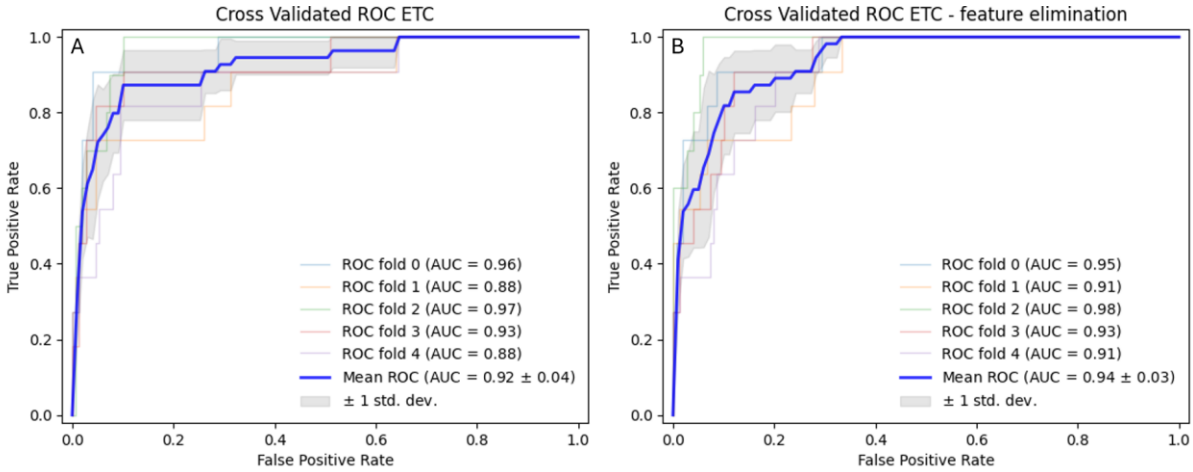


Figure 7 ROC curves of the ETC model for detection of oCAD on five stratified folds of the training set, before (A) and after (B) feature selection.

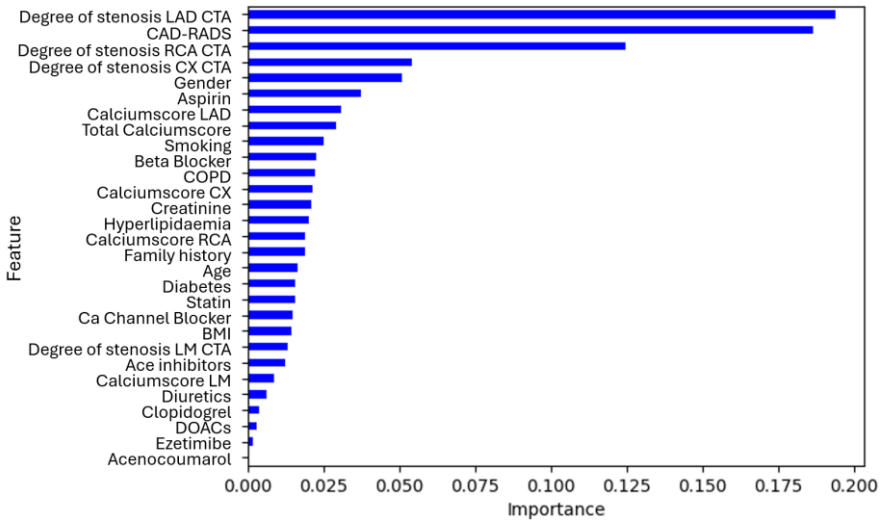


Figure 8 Ranked feature importance of the ETC model before feature selection.

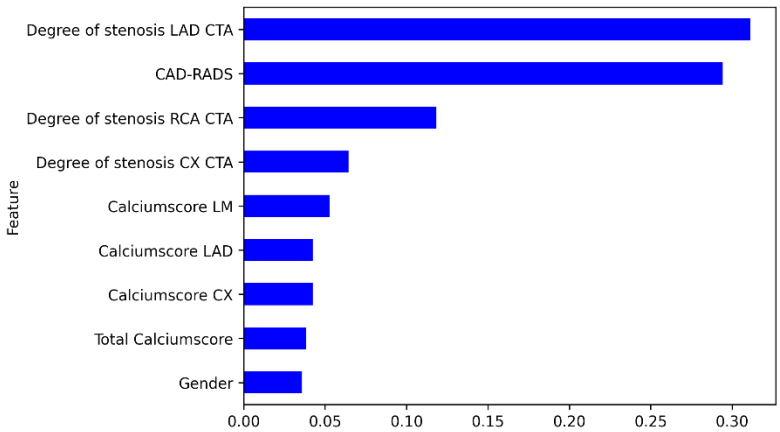


Figure 9 Ranked feature importance of the ETC model after Boruta feature selection.

Nine features were selected using Boruta feature elimination in three or more of the folds and were included in the feature selection. Degree of stenosis in the LAD, followed by CAD-RADS score, and degree of stenosis in the RCA and CX were ranked the most predictive features. Again, showing that CCTA features are most important to the predictive performance of the ETC model. Left main calciumscore outranks LAD, CX and total calciumscores. The final feature that is selected is gender. These features and their respective ranked feature importance can be seen in figure 9. Three features were selected in two or less folds, which were degree of stenosis in the LM, aspirin usage and hyperlipidaemia. This shows that they were informative in some folds. A more comprehensive dataset is needed to conclude their significance to information gain. The performance on the training set of the model after feature selection can be seen in figure 7. The mean ROC-AUC after feature selection was 0.94, which was not significantly improved compared to 0.92 before feature selection (p = 0.57). However, the performance of the model after feature selection achieved a lower standard deviation, suggesting that the model suffered less from the impact of noise.

3.3.3 Model evaluation and comparison to expert readers on the test set

The model achieved an accuracy of 0.93, precision of 0.50, recall of 0.94 and an ROC-AUC of 0.96 using the validation set as shown in figure 10, which was not significantly different compared to the performance on the training set (p = 0.53). The expert readers achieved an accuracy of 0.94, precision of 0.54 and recall 0.83. The performance of the expert readers has been plotted into the ROC curve of the Extra Trees Classifier, including 95% confidence intervals for sensitivity and specificity. McNemar Chi-Squared test indicated that there is no statistically significant difference between the performance of the model and that of the expert readers (p = >0.99).

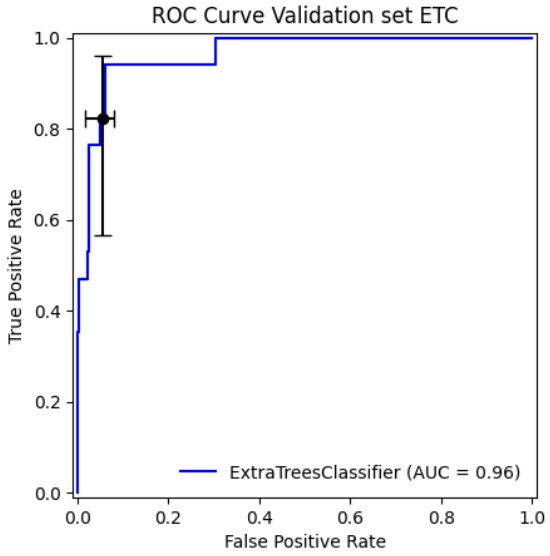


Figure 10 Performance of the ETC model after feature selection on the test set. The performance of the expert readers is plotted in black, with corresponding 95% confidence intervals for sensitivity and specificity.

3.4 Discussion

In this study we have developed and validated a machine learning model to diagnose obstructive CAD in patients without prior history of CAD, based on clinical data, medication usage, and imaging data from CACS and CCTA. Our results indicate that the Extra Trees Classifier (ETC) achieves an ROC-AUC of 0.96, showing a comparable performance with expert readers in predicting oCAD ($p > 0.99$). The impact of most non-imaging features on the model's predictive capacity is shown to be limited, except gender. This study highlights the potential of machine learning in oCAD diagnosis.

To our knowledge, this is the first paper to combine CACS and CCTA features in the prediction of oCAD using machine learning. Mohamed *et al.* have shown the added value of the combination of CACS and CCTA for the diagnosis of CAD using a statistical model, which achieved an AUC of 0.87 [51]. Motwani *et al.* combined CACS and CCTA features in a ML model in their prediction of 5-year all cause mortality for patients with suspected CAD, AUC 0.79 [52]. Even though, the focus of Motwani *et al.* is different compared to this study, they have also found imaging features to be the most important (17 out of top 20 features).

Baskaran *et al.* used CCTA features and patient characteristics to predict revascularization within 1 year using an XGBoost model [33]. They achieved an AUC of 0.941, which is comparable to the 0.96 found in this paper. However, the percentage of oCAD in their study was almost twice as high compared to our study, 13% vs 6.8%, respectively. This might have boosted their positive prediction; however, no information was provided about their achieved precision.

Multiple papers have used clinical data combined with CACS features to predict oCAD with machine learning. Aref *et al.,* Ren *et al.,* and Glowacki *et al.* achieved high negative predictive values, ranging from 94 to 100% which is to be expected since a low CACS is known to be an excellent negative predictor of oCAD [17,53–55]. However, they struggled with their precision with precision values ranging from 38 to 43%. This study achieved a precision of 50% with an oCAD percentage of just 6.8%, compared to the 16 to 18% found in these studies, which shows the added value of combining CCTA and CACS as Mohamed *et al.* have shown as well [51]. Similar to what was found in our study, they all found that imaging features were most important in the prediction of their models. Interestingly, Aref *et al.,* Ren *et al.,* Glowacki *et al.,* and Baskaran *et al.* found that age was also a high-ranking predictor, which was not found in our study [33,53–55].

This study has several limitations. First, we analysed data from a single centre, which makes generalizability to other hospitals not straight forward. The inclusion of this study was quite unique, since only patients referred by general practitioners were included. However, because most features are standardized, it is possible for other centres to retrain and validate this machine learning model on their data. Centre- and population-specific optimized hyperparameters will yield the best predictive performance for other centres.

Secondly, it is important to address the characteristics of our dataset. Because general practitioners referred the patients, they can be classified as relatively low-risk patients. This results in an imbalance in the dataset, as the number of patients who were classified as having oCAD was only 6.8%. Furthermore, even after extensive scrutiny our dataset might still contain some imperfections. Risk-factors were attained through input of the patient, which are prone to misunderstandings about for instance family history of cardiovascular disease. Also, the dataset contains possible confounders with the outcome. Since pharmacological management is the primary treatment for suspected CAD, some medication prescriptions can be seen as confounders to oCAD [56]. Because the expert readers also have access to these variables, it was chosen to keep them in the model. The effect of these

imperfections and possible confounders is limited since these features were removed during feature elimination.

Thirdly, it was chosen not to normalize the data, as not to suffer from information loss as a result of loss of range of variables. Given that the goal of this study is to incorporate the developed model into a tool to support physicians' decision-making, it was decided to keep data input as simple as possible. This approach aims to simplify the development of the tool. However, the predictive performance of mainly CACS features might have been enhanced if they were normalized with log scaling, since the distribution appears to confirm to the power law [57]. Normalization of these features might have improved the predictive performance of the model slightly.

Fourthly, the performance of the model was evaluated using a combination of precision, recall, accuracy, and ROC-AUC. The model achieved high recall, accuracy, and AUC, but a precision of 0.50. This can be explained by the imbalance of the data and the study's focus. The dataset contained few cases from which to learn positive predictions, requiring more data to improve positive predictive performance. The model was deliberately trained to favour false positives over false negatives, as a false negative can have severe negative consequences for the patient. Although a precision of 50% may seem like a 'coin flip,' it is comparable to the precision of our expert readers, who achieved a precision of 54% on the same data. In comparison to the diagnostic yield of 37,6% of oCAD in elective ICA procedures observed in large multi-center study, with over 35000 patients, our achieved precision is distinctly higher [53].

Finally, we classified patients as having oCAD if follow-up included a conclusive ICA for oCAD. For the assessment of performance of the expert readers a conclusive ICA procedure within 120 days after CCTA was seen as a correct prediction. The period was altered from the standard 90 days, since the inclusion of the data covered the COVID period [49,52]. This resulted in three more correctly predicted patients for the expert readers and a slightly increased reference for the model’s performance. For further research, not covering the COVID period, 90 days should be sufficient. Using a conclusive ICA as definition for oCAD, means we might have missed patients who were not referred for ICA. This bias might lead to underestimation of patients with oCAD. Two falsely positive predicted patients by the model were not referred for ICA but have confirmed ischemia with perfusion imaging. This indicates that the dataset contains additional oCAD cases that have already been accurately predicted by the model, without ICA confirmed oCAD diagnosis. This suggest that the precision of the model might actually be 5 to 10% higher.

3.5 Conclusion

We have developed and validated a machine learning model to diagnose obstructive CAD in patients without prior history of CAD, based on clinical data, medication usage, and imaging data from CACS and CCTA. Our findings demonstrate that the Extra Trees Classifier performs comparable to expert readers in predicting oCAD. The impact of most non-imaging features on the model's predictive capacity is shown to be limited, except gender. The model can assist expert readers in their decision-making by serving as a risk stratification tool. This study highlights the potential of machine learning in oCAD diagnosis.

4. Future perspectives

The model can serve as a risk stratification tool to assist expert readers in their decision-making. It will be interesting to identify the shared predictive performance of model assisted decision-making by expert readers. The model has shown to predict cases differently, offering new insights to expert readers. Utilizing the model might reduce the number of false positives. Expert readers recorded 13 false positives, while the model recorded 16. If decision-making were based solely on cases that both predicted positive this number could be reduced to just 6, more than halving the amount of false positive predictions. However, this method would also reduce the number of true positives from 16 to 14 (expert readers 14, model 16 out of 17). Future research into the added value of model assisted decision-making will determine the predictive performance.

Testing the model on a new validation set is a next step to validate the model's predictive performance. New data is already available, as new patients have been scanned through the diagnostic pathway following the closure of the inclusion of our study. Although there is not much long-term follow-up, however most interventions occur within a short follow-up period. Comparing the model's performance to that of expert readers might be challenging due to the short follow-up, as there is not enough time for false negatives to develop. Additionally, it would be interesting to test the model on a different group of patients, such as those with higher risk (e.g., patients referred by cardiologists who inherently have more risk). There is plenty of local data available to validate the model on similar or slightly different populations, which will provide further insights into the model's performance.

Furthermore, better documentation of symptoms is crucial. Symptoms have not been included in this research due to poor documentation. The type of symptoms has shown significance in predictive performance for oCAD [33]. Accurate documentation of risk factors is equally important. Currently, this is mostly done by patients and is prone to errors. It would be beneficial for cardiologists and general practitioners to document this information to ensure precision and reliability. Better documentation might increase the prognostic value of risk factors.

Additionally, developing separate models for males and females could lead to even better predictive performance. Since gender is a variable that can be easily split, creating gender-specific models might enhance the prediction of oCAD. This approach acknowledges that women often present with different symptoms, and thus, a tailored model might further improve predictive performance for each gender [58].

Finally, the future of CCTA in risk stratification for oCAD looks promising. The implementation of plaque characterization and the identification of high-risk plaques will enhance diagnostics [59]. These advancements are expected to be included in CAD-RADS 3.0 guidelines and subsequently implemented in hospitals such as Isala. Looking further ahead, computer vision, if properly validated and implemented, could boost CCTA's diagnostic performance [60]. For instance, radiomics could add value through additions of quantitative information about plaques that goes beyond pure visual assessment, providing deeper insights into plaque characteristics [61].

References

1. Byrne RA, Rossello X, Coughlan JJ, et al. 2023 ESC Guidelines for the management of acute coronary syndromes. *Eur Heart J*. August 25, 2023. doi:10.1093/eurheartj/ehad191

2. Knoop Y, Wimmers RH, Vaartjes I, Bots ML. Cijferboek hartstichting hart-vaatziekten Nederland 2021. *Hartstichting Den Haag*. 2021:1-100.

3. StatLine - Overledenen; belangrijke doodsoorzaken (korte lijst), regio. Statline. Accessed September 25, 2023. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80202ned/table?ts=1663663653426

4. Gopalan C, Kirk E. Atherosclerosis. *Biology of Cardiovascular and Metabolic Diseases*. January 1, 2022:85-101. doi:10.1016/B978-0-12-823421-1.00002-0

5. Neumann FJ, Sechtem U, Banning AP, et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. *Eur Heart J*. 2020;41(3):407-477. doi:10.1093/eurheartj/ehz425

6. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). *Eur Heart J*. 2019;40(3):237-269. doi:10.1093/EURHEARTJ/EHY462

7. Virani SS, Newby LK, Arnold S V., et al. 2023 AHA/ACC/ACCP/ASPC/NLA/PCNA Guideline for the Management of Patients With Chronic Coronary Disease: A Report of the American Heart Association/American College of Cardiology Joint Committee on Clinical Practice Guidelines. *Circulation*. 2023;148(9):e9-e119. doi:10.1161/CIR.0000000000001168

8. Saboury B, Bradshaw T, Boellaard R, et al. Artificial Intelligence in Nuclear Medicine: Opportunities, Challenges, and Responsibilities Toward a Trustworthy Ecosystem. *Journal of Nuclear Medicine*. 2023;64(2):188-196. doi:10.2967/jnumed.121.263703

9. Dey D, Slomka PJ, Leeson P, et al. Artificial Intelligence in Cardiovascular Imaging: JACC State-of-the-Art Review. *J Am Coll Cardiol*. 2019;73(11):1317-1335. doi:10.1016/j.jacc.2018.12.054

10. Shahjehan RD, Bhutta BS. Coronary Artery Disease. *Medical Nursing*. August 17, 2023:110-114. doi:10.4038/jccp.v54i2.8031

11. Jebari-Benslaiman S, Galicia-García U, Larrea-Sebal A, et al. Pathophysiology of Atherosclerosis. *Int J Mol Sci*. 2022;23(6). doi:10.3390/ijms23063346

12. Hermida N, Balligand JL. Low-density lipoprotein-cholesterol-induced endothelial dysfunction and oxidative stress: The role of statins. *Antioxid Redox Signal*. 2014;20(8):1216-1237. doi:10.1089/ars.2013.5537

13. Kapustin AN, Chatrou MLL, Drozdov I, et al. Vascular smooth muscle cell calcification is mediated by regulated exosome secretion. *Circ Res*. 2015;116(8):1312-1323. doi:10.1161/CIRCRESAHA.116.305012

14. Aung SSM, Roongsritong C. A Closer Look at the HEART Score. *Cardiol Res*. 2022;13(5):255-263. doi:10.14740/cr1432

15. Winther S, Schmidt SE, Foldyna B, et al. Coronary Calcium Scoring Improves Risk Prediction in Patients With Suspected Obstructive Coronary Artery Disease. *J Am Coll Cardiol*. 2022;80(21):1965-1977. doi:10.1016/j.jacc.2022.08.805

16. Shreya D, Zamora DI, Patel GS, et al. Coronary Artery Calcium Score - A Reliable Indicator of Coronary Artery Disease? *Cureus*. December 4, 2021. doi:10.7759/cureus.20149

17. Shiga Y, Morii J, Idemoto Y, et al. A Coronary Artery Calcium Score of Zero in Patients Who Have Undergone Coronary Computed Tomography Angiography Is Associated With Freedom From Major Adverse Cardiovascular Events. *J Clin Med Res*. 2020;12(10):662-667. doi:10.14740/jocmr4335

18. Dzaye O, Razavi AC, Dardari ZA, et al. Modeling the Recommended Age for Initiating Coronary Artery Calcium Testing Among At-Risk Young Adults. *J Am Coll Cardiol*. 2021;78(16):1573-1583. doi:10.1016/j.jacc.2021.08.019

19. Giusca S, Schütz M, Kronbach F, Wolf D, Nunninger P, Korosoglou G. Coronary computer tomography angiography in 2021—acquisition protocols, tips and tricks and heading beyond the possible. *Diagnostics*. 2021;11(6). doi:10.3390/diagnostics11061072

20. Nicol ED, Norgaard BL, Blanke P, et al. The Future of Cardiovascular Computed Tomography: Advanced Analytics and Clinical Insights. *JACC Cardiovasc Imaging*. 2019;12(6):1058-1072. doi:10.1016/j.jcmg.2018.11.037

21. Miller RJH, Eisenberg E, Friedman J, et al. Impact of heart rate on coronary computed tomographic angiography interpretability with a third-generation dual-source scanner. *Int J Cardiol*. 2019;295:42-47. doi:10.1016/J.IJCARD.2019.07.098

22. Cury RC, Leipsic J, Abbara S, et al. CAD-RADS™ 2.0 – 2022 Coronary Artery Disease-Reporting and Data System: An Expert Consensus Document of the Society of Cardiovascular Computed Tomography (SCCT), the American College of Cardiology (ACC), the American College of Radiology (ACR), and the North America Society of Cardiovascular Imaging (NASCI). *JACC Cardiovasc Imaging*. 2022;15(11):1974-2001. doi:10.1016/j.jcmg.2022.07.002

23. Skinner JS, Smeeth L, Kendall JM, Adams PC, Timmis A. NICE guidance. Chest pain of recent onset: Assessment and diagnosis of recent onset chest pain or discomfort of suspected cardiac origin. *Heart*. 2010;96(12):974-978. doi:10.1136/hrt.2009.190066

24. Koo BK. The present and future of fractional flow reserve. *Circulation Journal*. 2014;78(5):1048-1054. doi:10.1253/circj.CJ-14-0249

25. Demir OM, Mitomo S, Mangieri A, et al. Diagnostic Accuracy of Microcatheter Derived Fractional Flow Reserve. *American Journal of Cardiology*. 2019;124(2):183-189. doi:10.1016/j.amjcard.2019.04.038

26. Adjedj J, De Bruyne B, Floré V, et al. Significance of intermediate values of fractional flow reserve in patients with coronary artery disease. *Circulation*. 2016;133(5):502-508. doi:10.1161/CIRCULATIONAHA.115.018747

27. Song YY, Lu Y. Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatry*. 2015;27(2):130. doi:10.11919/J.ISSN.1002-0829.215044

28. Gini Impurity Measure— An intuitive explanation using python | Towards Data Science. Accessed June 26, 2024. https://towardsdatascience.com/gini-impurity-measure-dbd3878ead33

29. Gupta B, Uttarakhand P, Rawat IA. *Analysis of Various Decision Tree Algorithms for Classification in Data Mining*. Vol 163.; 2017.

30. Morán-Fernández L, Bólon-Canedo V, Alonso-Betanzos A. How important is data quality? Best classifiers vs best features. *Neurocomputing*. 2022;470:365-375. doi:10.1016/j.neucom.2021.05.107

31. XGBoost Parameters — xgboost 2.2.0-dev documentation. Accessed June 17, 2024. https://xgboost.readthedocs.io/en/latest/parameter.html

32. Anuradha P, David VK. Feature Selection and Prediction of Heart diseases using Gradient Boosting Algorithms. *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021*. March 25, 2021:711-717. doi:10.1109/ICAIS50930.2021.9395819

33. Baskaran L, Ying X, Xu Z, et al. Machine learning insight into the role of imaging and clinical variables for the prediction of obstructive coronary artery disease and revascularization: An exploratory analysis of the CONSERVE study. *PLoS One*. 2020;15(6):24. doi:10.1371/JOURNAL.PONE.0233791

34. Bulant CA, Boroni GA, Bass R, et al. Data-driven models for the prediction of coronary atherosclerotic plaque progression/regression. *Scientific Reports 2024 14:1*. 2024;14(1):1-19. doi:10.1038/s41598-024-51508-7

35. Schonlau M, Zou RY. The random forest algorithm for statistical learning. *Stata Journal*. 2020;20(1):3-29. doi:10.1177/1536867X20909688

36. Choi BG, Park JY, Rha SW, Noh YK. Pre-test probability for coronary artery disease in patients with chest pain based on machine learning techniques. *Int J Cardiol*. 2023;385:85-93. doi:10.1016/j.ijcard.2023.05.041

37. Samaras AD, Moustakidis S, Apostolopoulos ID, Papandrianos N, Papageorgiou E. Classification models for assessing coronary artery disease instances using clinical and biometric data: an explainable man-in-the-loop approach. *Scientific Reports 2023 13:1*. 2023;13(1):1-15. doi:10.1038/s41598-023-33500-9

38. Breiman L. *Random Forests*. Vol 45.; 2001.

39. Shaaque R, Mehmood A, Choi GS, Shafique R, Ullah S. Cardiovascular Disease Prediction System Using Extra Trees Classifier. September 16, 2019. doi:10.21203/RS.2.14454/V1

40. Louridi N, Douzi S, El Ouahidi B. Machine learning-based identification of patients with a cardiovascular defect. *J Big Data*. 2021;8(1). doi:10.1186/s40537-021-00524-9

41. ExtraTreesClassifier. How does ExtraTreesClassifier reduce… | by Naman Bhandari | Medium. Accessed June 26, 2024. https://medium.com/@namanbhandari/extratreesclassifier-8e7fc0502c7

42. Roshan V, Stewart JHM, Joseph R, Stewart HM. Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal*. 2022;15(4):531-538. doi:10.1002/SAM.11583

43. Liashchynskyi P, Liashchynskyi P. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. December 12, 2019. http://arxiv.org/abs/1912.06059

44. GridSearchCV — scikit-learn 1.5.0 documentation. Accessed June 27, 2024. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

45. Kursa MB, Jankowski A, Rudnicki WR. Boruta – A System for Feature Selection. *Fundam Inform*. 2010;101(4):271-285. doi:10.3233/FI-2010-288

46. ExtraTreesClassifier — scikit-learn 1.5.0 documentation. Accessed June 26, 2024. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html

47. Accuracy vs. Precision vs. Recall in Machine Learning | Encord. Encord. Published November 23, 2023. Accessed June 17, 2024. https://encord.com/blog/classification-metrics-accuracy-precision-recall/

48. Boiten HJ, Ekmen H, Zijlstra F, van Domburg RT, Schinkel AFL. Impact of Early Coronary Revascularization on Long-Term Outcomes in Patients With Myocardial Ischemia on Dobutamine Stress Echocardiography. *Am J Cardiol*. 2016;118(5):635-640. doi:10.1016/J.AMJCARD.2016.06.026

49. Nakazato R, Arsanjani R, Achenbach S, et al. Age-related risk of major adverse cardiac event risk and coronary artery disease extent and severity by coronary CT angiography: results from 15 187 patients from the International Multisite CONFIRM Study. *Eur Heart J Cardiovasc Imaging*. 2014;15(5):586-594. doi:10.1093/EHJCI/JET132

50. Sun X, Xu W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Process Lett*. 2014;21(11):1389-1393. doi:10.1109/LSP.2014.2337313

51. Mohamed M, Bosserdt M, Wieske V, et al. Combination of computed tomography angiography with coronary artery calcium score for improved diagnosis of coronary artery disease: a collaborative meta-analysis of stable chest pain patients referred for invasive coronary angiography. *Eur Radiol*. 2024;34(4):2426-2436. doi:10.1007/S00330-023-10223-Z/FIGURES/5

52. Motwani M, Dey D, Berman DS, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: A 5-year multicentre prospective registry analysis. *Eur Heart J*. 2017;38(7):500-507. doi:10.1093/eurheartj/ehw188

53. Al'Aref SJ, Maliakal G, Singh G, et al. Machine learning of clinical variables and coronary artery calcium scoring for the prediction of obstructive coronary artery disease on coronary computed tomography angiography: analysis from the CONFIRM registry. *Eur Heart J*. 2020;41(3):359-367. doi:10.1093/EURHEARTJ/EHZ565

54. Ren Y, Li Y, Pan W, Yin D, Du J. Predictive value of CAC score combined with clinical features for obstructive coronary heart disease on coronary computed tomography angiography: a machine learning method. *BMC Cardiovasc Disord*. 2022;22(1):1-8. doi:10.1186/S12872-022-03022-9/FIGURES/5

55. Głowacki J, Krysiński M, Czaja-Ziółkowska M, Wasilewski J. Machine Learning-based Algorithm Enables the Exclusion of Obstructive Coronary Artery Disease in the Patients Who Underwent Coronary Artery Calcium Scoring. *Acad Radiol*. 2020;27(10):1416-1421. doi:10.1016/J.ACRA.2019.11.016

56. Neumann FJ, Sechtem U, Banning AP, et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromesThe Task Force for the diagnosis and management of chronic coronary syndromes of the European Society of Cardiology (ESC). *Eur Heart J*. 2020;41(3):407-477. doi:10.1093/EURHEARTJ/EHZ425

57. Cleophas TJ, Zwinderman AH. Introduction to Machine Learning in Medicine Part Three. In: *Machine Learning in Medicine*. Springer Netherlands; 2013:1-9. doi:10.1007/978-94-007-7869-6_1

58. Garcia M, Mulvagh SL, Merz CNB, Buring JE, Manson JAE. Cardiovascular Disease in Women: Clinical Perspectives. *Circ Res*. 2016;118(8):1273. doi:10.1161/CIRCRESAHA.116.307547

59. Cao JJ, Shen L, Nguyen J, et al. Accuracy and limitation of plaque detection by coronary CTA: a section-to-section comparison with optical coherence tomography. *Scientific Reports 2023 13:1*. 2023;13(1):1-8. doi:10.1038/s41598-023-38675-9

60. Baeßler B, Götz M, Antoniades C, Heidenreich JF, Leiner T, Beer M. Artificial intelligence in coronary computed tomography angiography: Demands and solutions from a clinical perspective. *Front Cardiovasc Med*. 2023;10. doi:10.3389/FCVM.2023.1120361

61. Ayx I, Froelich MF, Baumann S, Papavassiliu T, Schoenberg SO. Radiomics in Cardiac Computed Tomography. *Diagnostics*. 2023;13(2). doi:10.3390/DIAGNOSTICS13020307

Appendix A

Model selection

This section will cover the process of selecting one classifier to eventually compare to the expert readers in its performance on the validation set. Three classifiers have been evaluated on the training set: XGBoost, RFC and ETC. All three classifiers have been developed and evaluated similarly. Random state one has been used to split the training and validation set. The hyperparameters resulting from Gridsearch and cross-validation are shown in table A-1.

Table A-1 Hyperparameters of the XGBoost, RFC, and ETC models

Hyperparameters XGBoost	Hyperparameters RFC	Hyperparameters ETC
scale_pos_weight = 13,7	class_weight = 'balanced_subsample'	class_weight = 'balanced_subsample'
colsample_bytree = 0.8	max_depth = 10	max_depth = 10
gamma= 0.5	max_features = 'log2'	max_features = 'sqrt'
L_lambda= 0.001	min_samples_split = 20	min_samples_split = 10
learning_rate= 0.1	min_samples_leaf = 2	min_samples_leaf = 1
max_depth= 8	n_estimators = 50	n_estimators = 200
subsample= 0.4	bootstrap = False	bootstrap = True
min_child_weight = 1		criterion = 'gini'

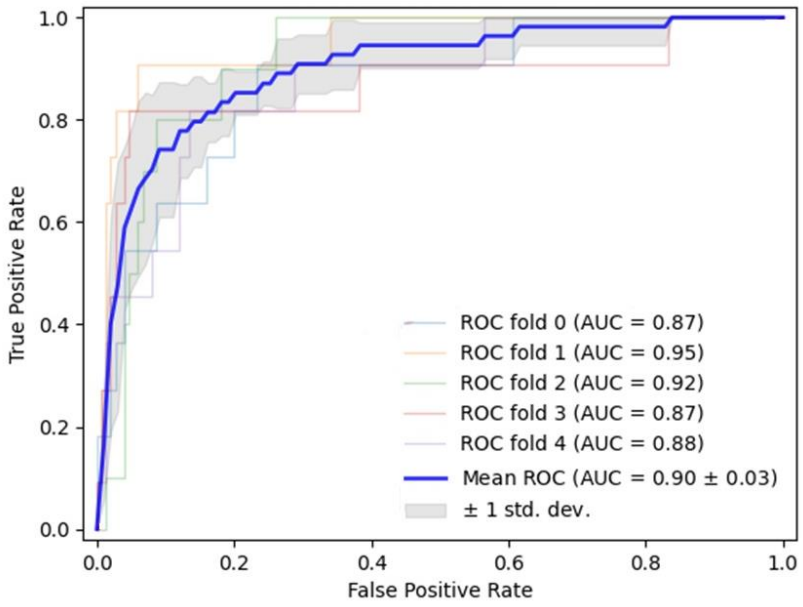


Figure A-1 ROC curves of the XGBoost model for detection of oCAD on five stratified folds of the training set. Mean ROC-AUC = 0.90

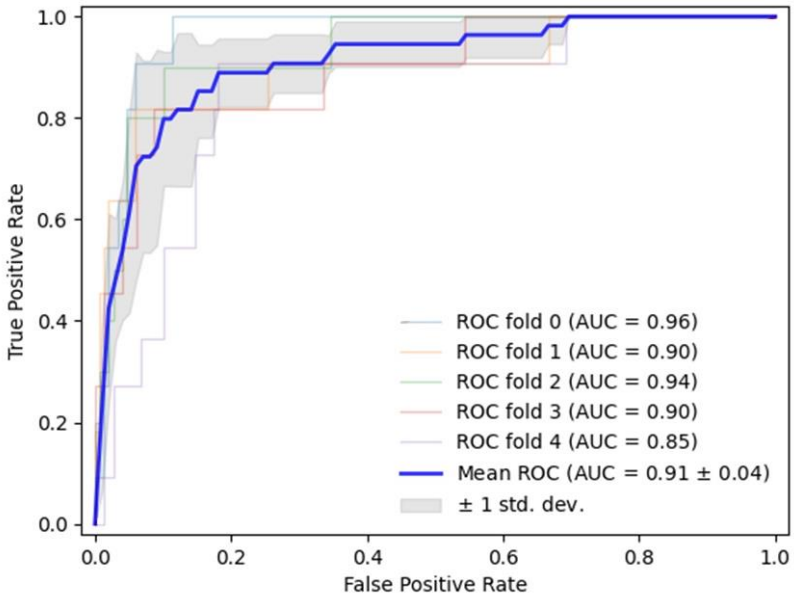


Figure A-2 ROC curves of the RFC model for detection of oCAD on five stratified folds of the training set. Mean ROC-AUC = 0.91

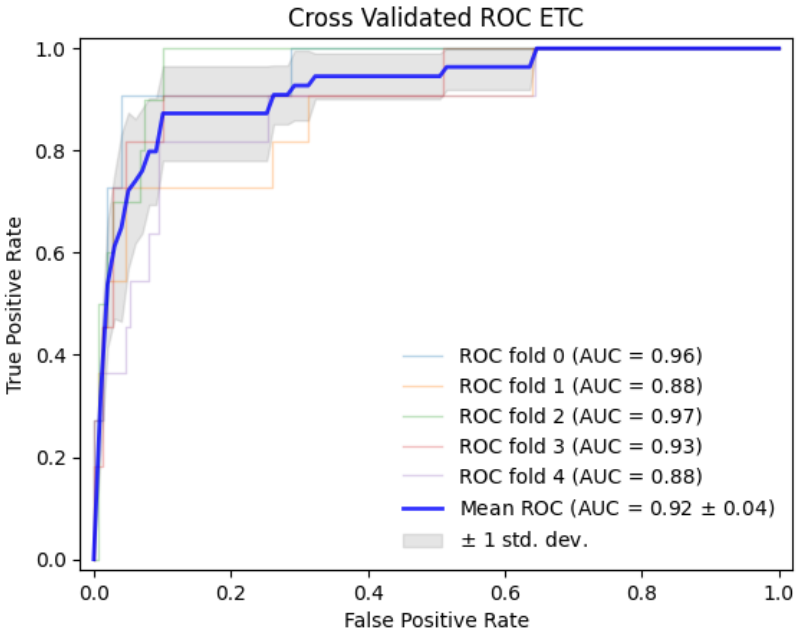


Figure A-3 ROC curves of the ETC model for detection of oCAD on five stratified folds of the training set. Mean ROC-AUC = 0.92

Table A-2 Performance metrics of the XGBoost, RFC, and ETC models on the training set without feature selection

	Accuracy	Precision	Recall	Mean ROC-AUC
XGBoost	0.9353	0.5385	0.5000	0.90 ± 0.03
RFC	0.9154	0.4400	0.7857	0.91 ± 0.04
ETC	0.9402	0.5500	0.7857	0.92 ± 0.04

The precision of the Random Forest and Extra Trees were better than the precision of the XGBoost model, as can be seen in table A-2. The XGBoost model produced a similar number of True positives, False positives, and False Negatives, where the others predicted more True positives compared to False negatives. The Extra Trees Classifier was preferred over the Random Forest Classifier because of its higher accuracy, precision, and ROC-AUC scores. The ROC-curves, shown in figure A-1 to A-3, show that the ETC achieved the highest mean ROC-AUC. However, since the standard deviations of the models overlap, the differences in mean ROC-curve were not expected to be significant. However, because the performance on all four performance metric was best for the Extra Trees Classifier, this classifier has been selected for further analysis. In chapter three of this thesis the ETC is further developed with feature selection, evaluated on the test set, and compared to the performance of expert readers.

Appendix B

Abstract for ICNC-CT 2024

Authors. JG Valk[1,2], PH Hiemstra[3], JD van Dijk2, JA van Dalen[4], RHJA Slart[5], CO Tan[6], M Mouden[7], BN Vendel[2]

1 Technical Medicine, University of Twente, Enschede, The Netherlands
2 Department of Nuclear Medicine, Isala hospital, Zwolle, The Netherlands
3 Research group IT-innovation in healthcare, Windesheim University of Applied Sciences, Zwolle, The Netherlands
4 Department of Medical Physics, Isala hospital, Zwolle, The Netherlands
5 Faculty of Science and Technology, University of Twente, Enschede, The Netherlands
6 Faculty of Electrical Engineering, Mathematics and Computer science, University of Twente, Enschede, The Netherlands
7 Department of Cardiology, Isala hospital, Zwolle, The Netherlands

Title. Applying machine learning to diagnose obstructive coronary artery disease using calcium scoring, coronary CT angiography, and clinical data.

Introduction. Traditional diagnostic risk assessment in patients undergoing CT-based calcium scoring (CACS) and coronary CT angiography (CCTA) is based on a limited number of clinical and imaging findings. Machine learning has demonstrated its suitability to use a greater number of features to model complex non-linear relations leading to a more accurate diagnostic performance. Our aim was to develop and validate a machine learning model to diagnose obstructive coronary artery disease (oCAD) in patients without prior history of CAD, using CACS, CCTA and clinical data.

Method. We retrospectively included 1254 patients without a prior history of CAD who underwent CACS and CCTA. Cardiac risk factors; smoking, hypertension, hypercholesterolemia, diabetes, family history of CAD; age; gender; body mass index; creatinine serum values; CCTA-outcomes, CACS and medication usage were registered at time of the scan. The entire dataset was split 4:1 into a training and test dataset, respectively. A XGBboost model was developed on the training set using fivefold stratified cross-validation and hyperparameter tuning. The test dataset was used to compare the diagnostic performance of the model to the performance of expert readers, looking at specificity, sensitivity, and ROC-analysis. The primary endpoint was oCAD on invasive coronary angiographies (ICA). Patients who were not referred for ICA were presumed not to have oCAD.

Results. A total of 85 (6.8%) of the 1254 patients (46% male, 60 ± 8 years of age) were diagnosed with oCAD. Age, total calcium score, body mass index, left anterior descending artery calcium score, and creatinine serum level were the highest-ranking features to predict oCAD. The performance of the XGBoost model did not differ from expert readers in predicting oCAD (p=0.84). The model and expert readers achieved a specificity of 0.97 vs 0.95, sensitivity of 0.71 vs 1.00, and an AUC of 0.96 vs 0.97, respectively.

Conclusion. The XGBoost model demonstrated a diagnostic performance comparable to that of expert readers. Features with negative predictive value in the diagnosis of oCAD were most valuable to the model such as high CACS. Currently, the model cannot replace expert readers, but it can contribute to the consistency of oCAD diagnoses and can be used to assist physician experts.