

The usability of interventional X-ray data for intraoperative prediction of coronary angiography procedure duration

Master thesis
Sietske Imming



The usability of interventional X-ray data for intraoperative prediction of coronary angiography procedure duration

by

Sietske Imming

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Monday December 14, 2020 at 10:00 AM.

Student number: 4269497

Supervisors: Dr. J.J. van den Dobbelsteen TU Delft
Prof. dr. B.H.W. Hendriks TU Delft, Philips Research

Thesis committee: Dr. J.J. van den Dobbelsteen TU Delft
Prof. dr. B.H.W. Hendriks TU Delft, Philips Research
Prof. dr. ir. J. Harlaar TU Delft

This thesis is confidential and cannot be made public until December 14, 2022.

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

Preface

Looking back at seven years of studying, I cannot begin to describe how precious my time in Delft has been and how thankful I am for everything it has given me. This is mostly because of the people around me and I would like to use this opportunity to express my gratitude.

After one year of Industrial Design Engineering, four years of Aerospace Engineering and a master in Biomedical Engineering, I can safely say I've explored it all and have really found a field I thoroughly enjoy. During my master, I had the chance to combine my passion for healthcare with my interest in technology. Also, I was able to finish the degree by writing my thesis on a very interesting and relevant project and got to fulfil my wish of learning about data analysis and machine learning. I would like to thank John for introducing me to the project and supervising me while giving me the freedom to explore different fields of interest. I also want to thank Benno, for all the useful advice and new insights that encouraged me to look further and expand my knowledge. Additionally, I want to thank Sjors for the help with my questions on machine learning, as well as Teddy for helping me to understand the medical side of the project. Furthermore, I would like to thank Ernesto for informing me about the ins and outs of the procedure and the cath lab team for showing me around. Last but not least, a big shout-out to you Joe, for introducing me to the field of machine learning and random forest and for always being there to help me with my questions and doubts.

Back in 2013, I arrived in the Netherlands as an international teenager from Portugal and hardly felt at home in this country that was so familiar yet new. Thanks to my lovely housemates, my grandma, many study sessions and many many parties and escalation nights, the experience turned into the best time of my life. I've made lifelong friends, lived in two cities I can now call home and was lucky enough to experience two extremely valuable exchange periods in Suriname and Brazil. None of this would have been the same if it wasn't for the entire JVB family; Luc, Marc, Joe, Flavie, Irene, Mick, Tamy and the rest of the crew. Thank you Maike and Laura, for making Huize Hendricks an equally warm, comfortable and incredible home. Ien, you're amazing. To my famalam; thank you, thank you, thank you. I am so grateful for all your love and support and I couldn't have wished for a better upbringing. No matter the distance, you've always felt close. Finally, thank you Laurent, for your infinite positivity and encouraging thoughts. You inspire me. This report marks the end of an era, but I know that many good times still lie ahead. I look forward to what is yet to come.

Sietske Imming
The Hague, December 2020

Abstract

Maximisation of treatment efficiency in hospitals can lead to significant growth in terms of patient satisfaction, staff productivity and hospital revenue. Scheduling for the operating department is currently done manually, using standard values for duration that vary with procedure type. This implies that no patient characteristics nor historical data are used to personalise predicted surgical duration. Furthermore, schedules are not updated when procedures are delayed. A relevant step towards dynamic scheduling is the realisation of real-time analysis of surgical workflow, based on intraoperatively acquired data. As of yet, the amount research performed into the use of automatically generated online data for duration prediction is limited. This research project involves an analysis of the usability and characteristics of interventional X-ray data, for predicting the total duration of coronary angiograms intraoperatively. A random forest classification algorithm is used to analyse each acquisition within the dataset and classify the total duration of the corresponding procedure as being below 10 minutes, 10 to 20 minutes, 20 to 30 minutes or 30 minutes and over. From the original dataset, 30 features were selected and implemented by the random forest model. An additional 22 features were generated to add data from prior acquisitions of the same patient. Recursive feature elimination was used to determine the final feature set.

Based on out-of-bag validation, the overall accuracy of the classification model was found to be 92.8%. Procedures with a duration lower than 10 minutes were correctly classified in 98.1% of the time. The classification accuracy for the other three classes ranges from 87.4% to 89.5%. Considering an average procedure duration of 11 minutes, the interventional X-ray data shows exceptional capabilities of classifying both standard and delayed procedures. For a deeper understanding of the model's abilities, the change in predicted duration class was evaluated with respect to procedural progress. This analysis shows that some Class 1 acquisitions are overestimated to be Class 2 acquisitions, but that overestimation of procedure duration rarely occurs for the other three phases. This implies that a prediction made beyond 10 minutes into the procedure can be perceived as the absolute minimum duration class of the procedure. Additionally, Class 3 and Class 4 predictions are found to always correspond to the minimum procedure time, independent of procedural progress. Class 2 and Class 3 acquisitions are underestimated up until 10 minutes and 20 minutes into the procedure, respectively. Class 4 acquisitions are correctly classified at a relatively earlier point with respect to progress and are never incorrect beyond 20 minutes.

To gain further knowledge on the implementability of the model, several procedures were analysed individually. New random forest models were generated using all data except for the acquisitions from the analysed procedure. As a result, the classification abilities for some procedures appeared to be different compared to prior analysis. A possible reason is that acquisitions from the same patient are dependent and cannot be split into training data and test data, as this might bias the results. Duration predictions for Class 1 continue to show overestimation and fluctuating class predictions. Nevertheless, the model still shows that it is capable of detecting delay at an early moment in the procedure, even when applied to entirely new data from unseen procedures. Given the fact that overestimation does not occur beyond 10 minutes, nor when Class 3 or Class 4 is predicted, the first prediction of a new duration class is concluded to be a reliable reference point. Further investigation has shown that most Class 3 and Class 4 detection occur within the first five minutes of a procedure. Therefore, the model is successful at predicting a total duration of 20 minutes or more at an early stage of the procedure. This could significantly benefit the hospital in terms of procedure planning and knowing when to request the next patient. In terms of prediction features, the acquisition frequency, longitudinal position of the operating table, cumulative procedure duration, patient age and cumulative cine acquisition time were found to be the most important. The implementation of these features in further research on CAG phases is recommended. Movement of the operating table seems to be particularly informative for workflow analysis. In terms of clinical application, further optimisation is required in order to enable accurate prediction for the duration of shorter procedures. Furthermore, the exact effect of the model on procedure scheduling requires validation by means of additional data such as existing schedules. Nevertheless, the model provides an accurate tool for the real-time monitoring of procedural workflow and detecting significant delay, which proves that the usability of the data goes far beyond machine maintenance and service only.

Contents

Preface	ii
Abstract	iv
1 Introduction	1
2 Background	3
2.1 Clinical Setting	3
2.1.1 Coronary Angiography	3
2.1.2 The Catheterization Laboratory	5
2.2 The Workflow Project	6
2.2.1 Project Vision.	6
2.2.2 Project Status.	6
2.3 Machine Learning Application	7
2.3.1 Optional Methods	7
2.3.2 Decision Trees.	8
2.3.3 Random Forest	9
3 Data & Methodology	10
3.1 Data Collection	10
3.1.1 Interventional X-ray Data	10
3.1.2 Procedure Type Determination	11
3.1.3 Machine Derived Features	13
3.1.4 Additional Features.	15
3.1.5 Operative Phase Implementation	17
3.1.6 Data Framing	21
3.2 Classification Test Setup	23
3.2.1 Sample Distribution	23
3.2.2 Class Selection	24
3.2.3 Hyperparameter Tuning	25
3.3 Recursive Feature Elimination	30
3.4 Classification Model Overview	31
3.5 Model Application Setup.	32
3.6 Feature Analysis Setup.	33
4 Results & Discussion	34
4.1 Classification Accuracy	34
4.1.1 Out-of-bag Accuracy	34
4.1.2 Cross-validated Accuracy.	35
4.1.3 Classification Timing	36
4.1.4 Model Application	39
4.2 Interpretation of the Classification Results	45
4.3 Feature Importance.	49
4.4 Feature Analysis	50
4.4.1 Acquisition Frequency	50
4.4.2 Longitudinal Position of the Operating Table.	51
4.4.3 Cumulative Procedure Duration	52
4.4.4 Patient Age Range	54
4.4.5 Cumulative Cine Acquisition Time	55
4.5 Interpretation of the Feature Analysis.	56
4.6 Clinical Implementation of the Model	59

5	Conclusion & Recommendations	61
5.1	Conclusion	61
5.2	Recommendations	62
	Bibliography	64
A	Distribution of Acquisitions	68
B	Distribution of Daily CAG Procedures	69
C	Results of the RFE	70
D	Complete Set of Logged Features	72

List of Figures

2.1	The direction of blood circulation in the heart and the heart's main coronary arteries. . .	3
2.2	Coronary angiograms of the left coronary artery and right coronary artery [27].	4
2.3	Setup of the main components of a cath lab [40].	5
2.4	Example of a decision tree for the data in Table 2.1 [8].	8
3.1	The main components of the interventional X-ray system. Original image retrieved from [40].	13
3.2	The directional axes of particular logged features. Original image retrieved from [40]. . .	15
3.3	The right and left anterior oblique rotation [51].	18
3.4	The caudal and cranial angulation [51].	18
3.5	The detection rate of each coronary artery phase within the dataset.	19
3.6	The detection rate of each RCA phase with respect to progress.	20
3.7	The detection rate of each LCA phase with respect to progress.	21
3.8	The distribution of procedure duration within the dataset.	23
3.9	The distribution of procedure duration with respect to individual procedures and selected classes.	24
3.10	The effect of the number of trees on the out-of-bag classification accuracy.	26
3.11	The effect of the in-bag fraction on the out-of-bag classification accuracy.	27
3.12	The effect of the number of sampled features on the out-of-bag classification accuracy. . .	28
3.13	The effect of the minimum leaf size on out-of-bag classification accuracy.	29
3.14	The misclassification probability against the number of features, based on 10 RFE applications.	30
4.1	Confusion matrix of the predicted versus the true duration based on out-of-bag data. . .	34
4.2	Confusion matrix of the predicted versus the true duration based on test data.	35
4.3	Procedure duration predictions throughout all procedures that took 0-10 minutes.	36
4.4	Procedure duration predictions throughout all procedures that took 10-20 minutes.	37
4.5	Procedure duration predictions throughout all procedures that took 20-30 minutes.	37
4.6	Procedure duration predictions throughout all procedures that took over 30 minutes. . .	38
4.7	All procedure duration predictions throughout procedure 4181.	40
4.8	All procedure duration predictions throughout procedure 4183.	40
4.9	All procedure duration predictions throughout procedure 4184.	41
4.10	All renewed procedure duration predictions throughout procedure 4184.	41
4.11	All procedure duration predictions throughout procedure 3221.	42
4.12	All procedure duration predictions throughout procedure 3124.	42
4.13	The procedure times at which the first predictions for Class 1 and Class 2 are made. . . .	43
4.14	The procedure times at which the first predictions for Class 3 and Class 4 are made. . . .	43
4.15	The out-of-bag importance of all features for the random forest classification model. . . .	49
4.16	The distribution of all values for acquisition frequency within the dataset.	50
4.17	The values for acquisition frequency with respect to progress and duration class.	51
4.18	The distribution of all values for longitudinal table position within the dataset.	51
4.19	The values for longitudinal table position with respect to progress and duration class. . .	52
4.20	The distribution of all values for longitudinal table position within the dataset.	53
4.21	The values for cumulative procedure duration with respect to progress and duration class. .	53
4.22	The distribution of the age of all patients within the dataset.	54
4.23	The distribution of all values for patient age range within the dataset.	55
4.24	The values for cumulative procedure duration with respect to progress and duration class. .	55
A.1	The distribution of acquisitions with respect to procedural progress.	68
B.1	A distribution of the number of CAG procedures per day.	69

List of Tables

2.1	Input data for the example decision tree.	8
3.1	The selected application names for CAGs and non-CAGs	12
3.2	The selected procedure names for CAGs and non-CAGs	12
3.3	The original features logged by the interventional X-ray system.	14
3.4	The extra features added to the interventional X-ray dataset.	16
3.5	The general angulation and rotation values of each operational step [28].	18
3.6	The selected classes for procedure duration and occurrence within the dataset.	24
3.7	The optimal feature set based on results from the RFE.	31
3.8	An overview of the hyperparameters of the final classification model.	31
3.9	The mean duration of the main phases of a CAG [28].	32
4.1	All CAGs that took place in the cath lab of on February 25th 2019.	39
C.1	The ranking of features based on the recursive feature elimination.	70
D.1	The original features logged by the interventional X-ray system.	72

Acronyms

ANOVA Analysis of variance.

AO Aorta.

CAG Coronary angiogram.

CAU Caudal.

CRA Cranial.

DAP Dose Area Product.

DSA Digital Subtraction Angiography.

DTW Dynamic Time Warping.

GUI Graphical user interface.

IBF In-bag fraction.

IE Information entropy.

IG Information gain.

LA Left atrium.

LAO Left Anterior Oblique.

LV Left ventricle.

LCA Left coronary artery.

OOB Out-of-bag.

PA Pulmonary artery.

RA Right atrium.

RAO Right Anterior Oblique.

RCA Right coronary artery.

RF Random forest.

RFE Recursive feature elimination.

RV Right ventricle.

SID Source-to-image distance.

SVM Support vector machines.

1. Introduction

Maximising treatment efficiency in a hospital can have a significant impact on patient satisfaction and hospital revenue. Rising healthcare costs cause it to be a relevant and popular research topic [23]. The increase in costs can be linked to the continuous development of advanced medical technologies, as well as an increased prevalence of diseases worldwide [28]. In order to deal with the increasing costs and demand for medical intervention, changes must be made in all departments of the hospital. One of the most expensive and resource-intensive areas of a hospital is the operating department [19]. Research into optimisation within this department would not only benefit the surgical staff and patients, but also largely affect the income of the entire hospital.

One way to increase efficiency in the operating department is by using complex planning procedures to obtain a surgical schedule that is realistic and accurate. Surgical delay does not only impact the current procedure, but also affects the schedule for the rest of the day. Realistic schedules can be created by obtaining insight into the workflow of individual procedures. Accurate, real-time predictions of the total procedure duration, including any delay, would increase the quality of such schedules substantially. A specific type of operating room that could be an appropriate starting point for intraoperative duration prediction is the cath lab, short for catheterization laboratory. This room allows for a limited number of procedures with repetitive, detectable steps and is used to diagnose and treat cardiovascular diseases. A structured use of the intervention room and its staff would save significant costs [50]. Currently, many catheterization laboratories experience down times and non-productive times. Improvements in this area could improve patient flow, safety and satisfaction, while putting an end to wasting equipment and incidental overtime [26].

A procedure that is regularly executed inside the cath lab is the coronary angiogram. The procedure is becoming increasingly common, due to the rising prevalence of cardiovascular diseases such as coronary heart disease [15, 28, 57]. Coronary heart disease is the leading cause of heart attacks, which occur when coronary arteries are blocked and blood can no longer flow to and from the heart freely. Coronary angiography is used to detect and localise such obstructions in the arteries. The cath lab is among the most important and complex places in the hospital [2, 50]. It is a room that can have high levels of uncertainty, stress and volume [26]. Part of its complexity is due to the fast pace of the procedures, as well as the large variety of equipment. Examples of the required equipment include the interventional X-ray system and the C-arm, health monitoring systems, intravascular imaging tools and visualisation workstations [50]. The equipment keeps track of the procedure by means of a large variety of parameters. This means that the data from the equipment might be able to provide useful information regarding the procedure and its progress. In order to test this hypothesis and use the data to analyse and predict workflow or progress, machine learning can be applied. It can be used to detect certain patterns and extrapolate these to predict future instances.

The benefit of combining machine learning and online data is that predictions regarding workflow can be made intraoperatively, which allows for the creation of a dynamic surgery schedule rather than a static one. Extensive research has been done on the different types of models that could be used for surgery scheduling and intraoperative duration prediction [1, 7, 17, 34]. Models that can provide real-time predictions generally show promising results with respect to prediction accuracy and applicability. Nevertheless, most of these models are trained using data that is manually collected, for example by means of interpreting video footage from the operating room. Therefore, the amount of input information into the machine learning model is limited, as each data point must be validated by hand. The number of cases used in online models typically ranges from 1 to 80, which shows that there is a significant research gap regarding tests that use bigger sets of data [56]. Therefore, research into the application of machine learning using real-time data that is automatically collected throughout the procedure is highly relevant. Data from the medical equipment in the cath lab is a suitable choice, due to the generation of data at each acquisition, represented by numerous features that could be used to indicate progress and workflow.

The main aim of this report is to analyse the usability of data from the interventional X-ray system. The scope of the research is limited to data from the interventional X-ray system of the Reinier de Graaf hospital, provided by Philips Research. Furthermore, only data from coronary angiograms is used. The overall objective of this research is specified by the following research question:

"To what extent can data logged by interventional X-ray systems be used to intraoperatively predict the total procedure duration class of a coronary angiogram and how do the most significant features affect the analysis?"

In order to answer the above research question, a general guideline is provided by the following sub-questions:

1. *How can a random forest classification algorithm be used to estimate procedure duration and analyse feature importance of the available data?*
2. *Which additional features could be added to the the model in order to improve its predictive ability?*
3. *What are the optimal features and model parameters that lead to the highest possible classification accuracy?*
4. *How effectively can the final model classify the procedures duration to be <10 minutes, 10-20 minutes, 20-30 minutes or >30 minutes?*
5. *When are the predictions made throughout the procedure and how could this affect the analysis of real-life procedures?*
6. *Which features from the optimal feature set are most useful for determining the procedure duration class and how do they change with respect to duration?*

In terms of the structure of this report, background information about coronary angiography and the cath lab is given in Chapter 2. This chapter also explains the overarching project of this research, as well as the principle theory of the random forest machine learning algorithm. Chapter 3 explains the methods of extracting and creating the optimal input dataset for the model, as well as the steps taken to set up the classification model. Additionally, it explains the method applied to test model applicability and to analyse the impact of individual features from the dataset. The results of the classification model are presented in Chapter 4, including an analysis of the applicability and the behaviour of the most significant features. Finally, the research question is answered in Chapter 5, along with recommendations for future research on this topic.

2. Background

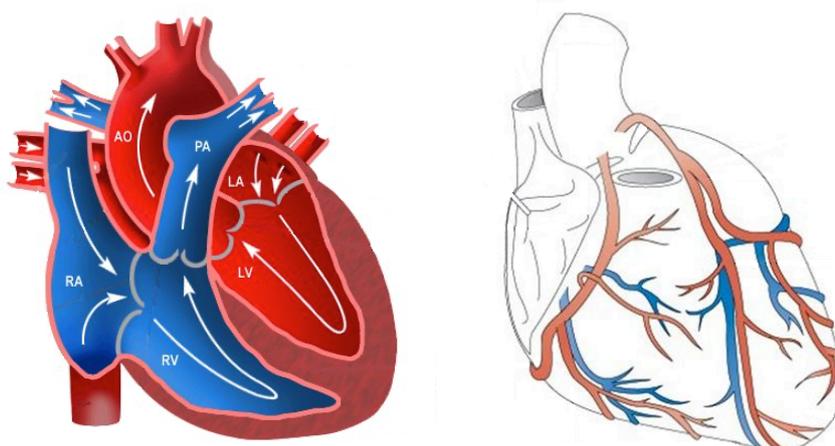
This chapter presents the background information that can be valuable for understanding the medical and technical aspects of this report. First of all, the clinical setting is presented in Section 2.1. This includes a description of the general structure and function of the heart and coronary angiography. Furthermore, the medical staff, equipment and operating room required for coronary angiography are explained, in order to provide a good understanding of the medical procedure that forms the essence of this project. Thereafter, the overarching project of this research is presented in Section 2.2. Finally, the theory behind the used machine learning algorithm is explained in Section 2.3.

2.1. Clinical Setting

The medical procedure to examine coronary vessels, the coronary angiogram, is described in Section 2.1.1. The specific type of operating room that is required is explained in Section 2.1.2.

2.1.1. Coronary Angiography

Within the branch of cardiology, coronary procedures specifically involve the arteries that surround and supply the heart [20]. The heart is part of the cardiovascular system, which is responsible for maintaining the body's homeostasis by making sure oxygen and nutrients are delivered to all cells in the body and waste products are removed from them [57]. The heart consists of four chambers, where the upper collecting chambers are called atria and the two lower pumping chambers are the left and right ventricles. A diagram of the heart can be seen in Figure 2.1a. Deoxygenated blood enters the heart on the right side, through the inferior and superior vena cava. The blood moves from the right atrium (RA) into the right ventricle (RV) and from there, it is pumped into the pulmonary artery (PA) which distributes it into the pulmonary circulatory system. When the blood reaches the alveoli of the lungs, CO₂ and water are exchanged with oxygen, leading to a significant increase in blood saturation. The pulmonary veins guide the oxygenated blood back to the heart, where it enters through the left atrium (LA). After the blood has entered the left ventricle (LV), it is pumped into the aorta (AO) and thereafter to the rest of the body, to supply it with oxygen and nutrients. After the diffusion of oxygenated blood into the tissue and the collection of waste products, the blood travels back through the veins towards the inferior and superior vena cava, to complete the cycle [57].



(a) The circulation of blood around the heart [52].

(b) The main coronary vessels [44].

Figure 2.1: The direction of blood circulation in the heart and the heart's main coronary arteries.

Angiography refers to the radiological imaging of the arteries and veins using X-rays, after being injected with a contrast agent. Coronary angiography specifically focuses on the heart and is used to diagnose cardiovascular problems [57]. For example, a coronary artery may show narrowing (stenosis) or blockage, reducing the oxygen supply to the heart. In this case, angiography can be used to detect the exact location, so that appropriate treatment can be decided upon. An example of a critical case of stenosis can be seen in Figure 2.2a. Angiography can be used to fix still images on film or capture moving images, where the process of generating moving images is referred to as fluoroscopy. In both applications, the resulting images can be viewed real-time by physicians. The use of a catheter to introduce the contrast agents into the vessels or heart chambers is called catheterization. For cardiac catheterizations, catheters are inserted by means of an injection in the wrist or in the groin, where entry through the wrist is the most common and generally preferred option, due to of lower complication rates and less recovery time [28, 29]. Generally, the right wrist is used, but when access to the artery or aorta arch cannot be attained, the physician may divert to using the left wrist or the groin. Once the catheter has reached the aorta, both the left coronary artery (LCA) and right coronary artery (RCA) are captured. For both steps, contrast agent is injected at the tip of the arteries. Once a coronary angiogram of both arteries is made and the medical issue has been localised, the catheter is retracted from the body.

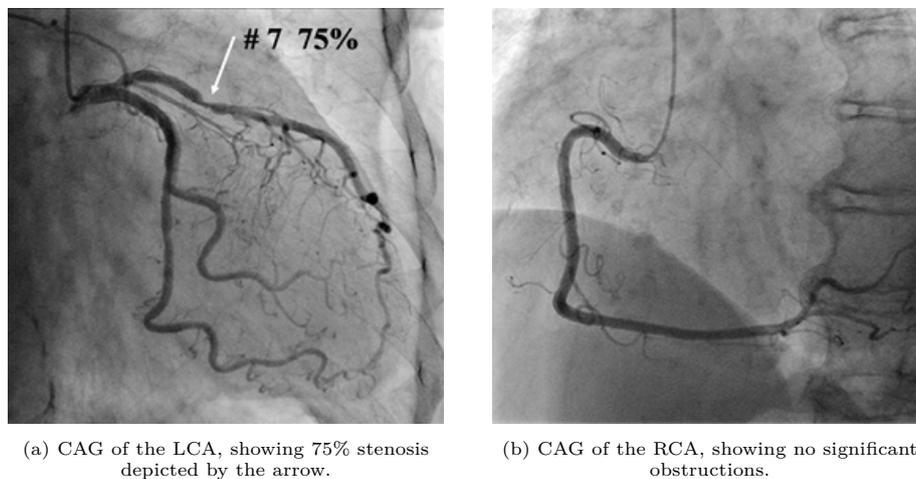


Figure 2.2: Coronary angiograms of the left coronary artery and right coronary artery [27].

The staff team required to perform a coronary angiogram generally consists of four members. The cardiologist is in charge and leads the procedure. Additionally, three support staff members are required, with the following division of roles:

- **Cardiologist** - The cardiologist carries the main responsibility during the procedure. He or she is in charge of the catheter and guides it to the desired location in the heart. Furthermore, the cardiologist is in control of the foot pedal, which is used to capture the fluoroscopic footage [28].
- **Scrub** - The scrub assistant assists with the handing over of instruments to the cardiologist. Furthermore, the injection of medication may be done by the scrub. The scrub handles the sterile equipment [57]. In academic hospitals, the presence of a cardiologist trainee can make the scrub redundant [38].
- **Circulator** - The circulator is responsible for the turnover, which refers to the preparation of the room before patient arrival and cleaning up afterwards. The circulator helps to transfer the patient to the table and prepares them for the intervention by attaching the ECG electrodes and inserting an peripheral intravenous line if necessary. The circulator handles the non-sterile equipment [57].
- **Monitor** - The monitor observes and records the patient's vital signs, as well as procedural measurements [57].

The Reinier de Graaf hospital assumes a procedure duration of 45 minutes when scheduling for the cath lab [31]. Due to its minimally invasive nature, two-thirds of all cath lab procedures are performed on outpatients. Outpatients are patients that do not have to stay overnight. For the outpatients, recovery time takes about 4 to 6 hours. Inpatients have to recover for approximately 90 minutes before returning to their rooms [29].

2.1.2. The Catheterization Laboratory

The cath lab is a special type of operating room that holds relatively modern and technically advanced medical equipment, suitable for minimally invasive surgery and interventional radiology. The use of radiological image guidance and minimally invasive access to vessels allows for very precise targeting, applicable to both diagnosis and treatment [28]. Typically, the cath lab is a fast-paced and procedural environment [46]. Even though interventional radiology has a wide range of applications, most procedures involve the heart. Coronary angiography is one of the most frequently executed procedures done inside the the cath lab [28].

The catheterization procedure requires three principal work zones. The first work zone is the preparation and recovery area for preoperative assessment, post-procedure care and recovery. The catheterization procedure itself takes place in the cath lab suite, which is a controlled environment where scrub attire and cover gowns must be worn by all clinical staff [29]. Finally, there is a control room for supervision of the procedure by means of imaging control panels and monitors [33]. Essential equipment inside the cath lab includes an interventional X-ray system, a mobile patient table, a control panel and a monitor [20]. The general setup is shown in Figure 2.3. The interventional X-ray system contains an image intensifier, used to improve image quality by means of amplifying the X-ray beams and converting the X-ray image into a light image. The light image can then be displayed on the monitor through an image processor [39].



Figure 2.3: Setup of the main components of a cath lab [40].

Due to its shape, as shown in the figure above, part of the X-ray system is referred to as the C-arm. The main benefit of the C-shape is that it allows for both orbital and angular movement, while maintaining the subject of interest at the isocenter and keeping X-ray scattering to a minimum. This also means that the patient can be examined from head to toe using a large range of angles, without requiring any movement from the patient. Furthermore, the X-ray system allows for three dimensional imaging, when a particular software program is used [20]. The specific interventional X-ray system used in the cath lab of the Reinier de Graaf hospital is the Allura Xper FD20 X-ray system [42]. The X-ray system logs each procedure by means of many different parameters. This is the data that makes up the main dataset used in this report, which is further elaborated upon in Section 3.1.1.

2.2. The Workflow Project

This section contains a concise summary of the overarching project that this report contributes to. Section 2.2.1 explains the relevance and vision of the project. Section 2.2.2 explains its status and future steps that will be taken to achieve its final goal.

2.2.1. Project Vision

This report is part of an overarching project, initiated through collaboration between Delft University of Technology, Philips Research, Philips Healthcare and the Reinier de Graaf hospital in Delft, aiming to create a smart environment that provides insight into the workflow in the cath lab. A thorough understanding of the organisation inside the room can help detect ineffectiveness and improve efficiency. A detailed analysis of procedural workflow would allow for a better understanding of the procedural steps, which in turn could lead to a higher predictive accuracy of the remaining procedure time. Various tools could be of use for this analysis. One of the selected devices for monitoring activity inside the cath lab is a camera, due to its high sensitivity. A total of five cameras are used to record movement inside the room from different angles. These could be used to observe the use of medical equipment and instruments by the staff and track general movement inside the cath lab. Other possible sources of information include the medical equipment inside the cath lab. The information can be used to detect patterns in the data and obtain insight into different aspects of the procedure. Certain patterns could be linked to procedural phases and provide information on the progress of the procedure. Furthermore, workload of each medical staff member could be analysed. The analyses can be used to detect inefficiencies that delay or complicate the process, providing the necessary background information for seeking solutions and overall improvement.

Generally, the remaining duration of a procedure is estimated by the physician in the room, based on a personal perspective. Surgical schedules are typically made using values for procedure duration established from historical data. It is uncommon to take personal details such as patient characteristics or the composition of the surgical team into account for this. Gaining insight into the separate phases of a procedure and the effect of the aforementioned parameters would benefit the scheduling process significantly. Moreover, being able to accurately predict remaining duration real-time would increase the efficiency of patient inflow and outflow, which could eliminate unnecessary turnover aspects such as waiting for the next patient to arrive. Furthermore, maximising the cath lab occupancy would significantly benefit the hospital from a financial aspect.

2.2.2. Project Status

The first steps of the project consist mainly of writing protocols to establish an agreement between the project owners and the hospital. General privacy regulations and medical policies result in strict rules that must be applied to the process of recording footage inside the cath lab. In August 2020, all the agreements were signed and the cameras were set up inside the cath lab. In October 2020, the first footage was recorded.

Steps in the nearby future of the project include the collection and analysis of data. An algorithm must be used to analyse the footage and possibly detect procedural phases automatically, using all the available and collected information. Additionally, information regarding the phases can be derived from other devices such as the interventional X-ray system. A desired outcome of the project would be to shift from a static to a dynamic schedule and to be able to analyse the workflow real-time and intraoperatively. In due course, the system might be applicable to other types of operating rooms and procedures. The aim of this report is to contribute to an analysis of the feasibility of using machine data to monitor workflow and improve procedural planning.

2.3. Machine Learning Application

This section discusses the principal theory behind the selected machine learning algorithm. Section 2.3.1 describes why a random forest algorithm was selected to be the most appropriate method. Section 2.3.2 explains the concept of a decision tree, which is the fundamental element of a random forest model. Section 2.3.3 elaborates on the random forest learning method itself.

2.3.1. Optional Methods

To be able to interpret large sets of data and use this information to predict procedure duration, machine learning can be applied. Machine learning is a "branch of artificial intelligence focused on building applications that learn from data and improve their accuracy over time, without being programmed to do so" [22]. The main advantage of using machine learning is that the model is able to learn automatically using the information available. This implies that the model can also be used to predict the duration of future procedures that are not part of the current dataset. Given the fact that the data from the interventional X-ray system can be acquired and used real-time, this also implies that the model can predict duration intraoperatively. Therefore, the predictions could contribute to dynamic procedure scheduling, making the general throughput of the cath lab more efficient.

There are various types of machine learning methods that would be appropriate for the purpose of intraoperatively predicting the total procedure duration and obtaining an insight into the most significant information within the dataset. The first option is linear regression, which uses the linear least-square method to algebraically fit a model that best represents the dataset. This is a model that is easy to interpret and the effect of each feature on the response variable can be observed from the regression equation. Furthermore, each coefficient in the regression equation represents an independent variable for which there are no limits, making it a suitable method for large datasets with many different features. However, the formula can become quite extensive and no information is provided on the correlation between features. Furthermore, all features are taken into account for the construction of the regression equation, meaning that no information is provided on whether some features are unnecessary or lead to overfitting. Additionally, the model assumes a linear relationship between the input variables and output, which is not always the case.

An alternative method would be to use support vector machines (SVM). This is a classification technique, but is relatively sensitive to noise and not computationally efficient for large datasets [56]. Other methods that are not appropriate for handling large datasets with many different features include Markov models and Dynamic Time Warping (DTW). Neural networks are also a possible solution, but this method also lacks computational efficiency when compared to some other methods. None of the aforementioned methods are capable of providing sufficient information regarding feature importance. Since the aim of this report is to understand the usability of the interventional X-ray data, rather than to merely apply it, it is desirable to use a machine learning technique that provides information on the extent to which each feature affects the predictions for procedure duration. A random forest algorithm, which consists of multiple decision trees, is a suitable technique to analyse feature importance, as it can assess the importance of particular features by excluding them from certain permutations and measuring the reduction in prediction accuracy. The exact theory behind the random forest method is described in the next two sections.

2.3.2. Decision Trees

A decision tree is a support tool for estimating an unknown property of the input data, by asking consecutive questions regarding known parameters. Each point that holds a new question is called a node and the connections between the nodes are called edges. Each node has one incoming edge. The combination of nodes and edges give the graph a tree-like shape, hence its name [9]. The input data for a decision tree consists of individual data points, where each data point is represented by multiple components called features. In vector format, this can be represented as $\mathbf{V} = (x_1, x_2, \dots, x_N)$, where x represents a feature and N is the total number of features. Table 2.1 provides an example of a dataset with three data points. Each data point represents a type of animal, characterised by three features.

Table 2.1: Input data for the example decision tree.

Sample	Features		
	Feathers	Ability to fly	Fins
1	Yes	Yes	No
2	No	No	Yes
3	Yes	No	No

A simple example of a decision tree to determine the type of animal is presented in Figure 2.4. According to the decision tree, animal 1 must be a hawk. As the tree is shallow and the number of features is limited, the chance of misclassification is relatively high. However, by increasing the number of features and data points used to make the decision tree, the classification accuracy can be increased.

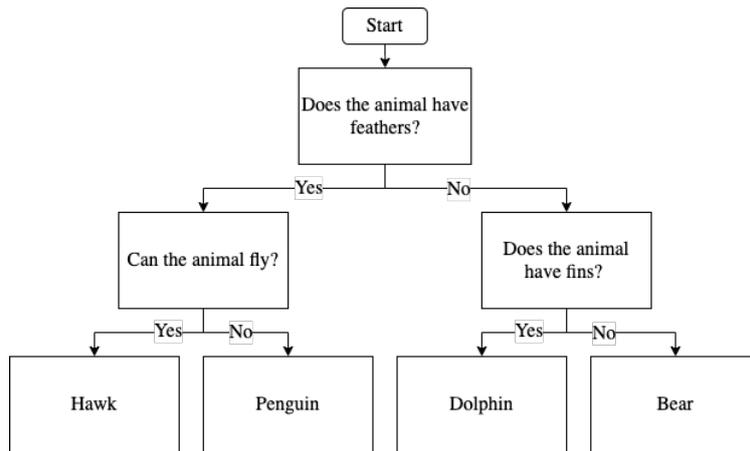


Figure 2.4: Example of a decision tree for the data in Table 2.1 [8].

Each internal node is split by two functions that determine the path followed by the datapoint. For simple datasets, these functions can be generated manually. For complex data, the model must evaluate the optimal functions for each node itself, based on training data. In a supervised task, training data refers to a set of data for which the response variable is already known. The response variable is the answer that one seeks to gain using the decision tree, such as the type of animal in the given example. Each data point, represented by different features and a response variable, is called a training point. In an unsupervised task, the training point contains no response variable and only consists of the feature responses. In this case, the model must detect clusters and specific patterns and define the output parameters by itself [9].

The optimal split functions are determined by analysing the amount of information gained from splitting the data in a certain way. This concept is referred to as maximisation of the Information Gain. In order to understand how Information Gain is deduced, the concept of Information Entropy must be explained, which refers to the degree of variance within the dataset. A dataset with many repetitive datapoints represents a low entropy, whereas many different datapoints result in a high entropy.

The Information Entropy of a particular dataset, where each datapoint could be assigned to a class C , is mathematically formulated as shown in Equation (2.1). In the example decision tree, the classes are represented by the leaf nodes.

$$E_{branch} = - \sum_i^C p_i \log_2 p_i \quad (2.1)$$

In Equation (2.1), p_i represents the probability that a data point belongs to class i . When the data is split by a test node, the remaining two subsets are called branches. The entropy of each branch is evaluated based on the number of datapoints from each class left after splitting, by means of Equation (2.1). When the two entropies are calculated, the quality of the split is determined by weighting the entropy against the total number of data points in each branch. This is shown by Equation (2.2), where lb refers to the left branch after the split and rb to the right branch.

$$E_{split} = p_{lb} E_{lb} + p_{rb} E_{rb} \quad (2.2)$$

Finally, the Information Gain (IG) is calculated as shown in Equation (2.3). The aim is to minimise the Information Entropy that would result from a particular split, thus to maximise information gain.

$$IG = 1 - E_{split} \quad (2.3)$$

Based on the Information Gain, the most informative features and its corresponding functions are assigned to the split. The process of splitting occurs until no further splits can be made or when the restrictions of the tree have been met. The terminal (leaf) node presents the most probable class or value, based on the presented node tests throughout the decision tree and the type of decision tree. Classification regression trees aim to classify input data, whereas regression decision trees are meant to predict continuous values [9, 21, 30]. In this report, classification trees are used to estimate procedure duration. This is because having fewer output options, with respect to regression, is likely to result in a higher prediction accuracy. This in turn increases the reliability of the analyses regarding individual features and data usability. Once the potential of classifying procedure duration and faulty aspects have been analysed, regression could be used to further analyse the predictive abilities of the interventional X-ray data. This is however beyond the scope of this research project.

2.3.3. Random Forest

A random forest is an ensemble of randomly trained decision trees. This implies that all trees that make up the forest are different, improving generalisation and robustness [4, 6]. There are various parameters that affect the generalisation of the model, the accuracy of the decision and computational efficiency. These parameters must be selected based on data characteristics such as size of the dataset and the presence of noise. More information on selecting such parameters, specifically applied to this report's model, can be found in Section 3.2.3.

A random forest uses a technique called Bootstrap Aggregation, commonly known as bagging [4]. This is a technique in which each decision tree is trained using a different subset of the original input dataset. The used observations are randomly selected with replacement, thus some observations may be selected multiple times whereas other do not contribute to the sample. The observations that are not used to construct a particular decision tree, are called out-of-bag observations [3]. These observations are unseen to this specific tree and can later on be used to test the accuracy of the tree. Given the fact that different subsets of training data are used for each tree and split nodes considers varying feature sets as well, a random forest shows a combination of bagging and randomised node optimisation. The use of varying training sets to construct each tree reduces the influence of dominant data points and noise, minimising overfitting and increasing the prediction accuracy when the model is exposed to new data [9, 56]. The concept of overfitting is further explained in Section 3.2.3.

3. Data & Methodology

This chapter describes the extraction and optimisation of input data for the classification model, as well as the required steps for building the model. The aim of Section 3.1 is to explain the interventional X-ray data, whereas Section 3.2 elaborates upon the test setup of the classification model. The combined optimisation of the dataset and model is explained in Section 3.3, followed by an overview of the final classification model in Section 3.4. The method for testing the applicability of the model is explained in Section 3.5. Finally, the setup for the analysis of individual features with respect to the procedure is presented in Section 3.6.

3.1. Data Collection

In this section, the methods used to extract and create the data are explained. Section 3.1.1 elaborates on the data from the interventional X-ray system. Section 3.1.2 explains the method used to filter CAG procedures from the data. Section 3.1.3 describes the features that are currently logged by the interventional X-ray system, whereas Section 3.1.4 presents additional features that could improve the accuracy of the model. Finally, Section 3.1.6 presents how the data is split into training and test data.

3.1.1. Interventional X-ray Data

The data used for this report is provided by Philips Research and is extracted from the interventional X-ray system in the cath lab of the Reinier de Graaf hospital. The intended purpose of this data is to provide a technical overview, to be used for machine maintenance or in case of machine malfunctioning. Each data point within the dataset represents one acquisition, which refers to the recording of a still or moving X-ray image. The process of generating live moving images is referred to as fluoroscopy, which is different from standard radiography, which generally fixes still images on film [24]. The vast majority of CAG acquisitions are fluoroscopic video clips, but there are occasional applications of single-shot recordings as well. Throughout the CAG procedure, acquisitions are executed to localise the catheter inside the body and to guide it towards the desired artery. Furthermore, acquisitions are made to record the vascular problem and its precise location. Different X-ray settings can be selected based on the desired purpose of the acquisition. A fluoroscopic acquisition is used for guiding the catheter and is generally not recorded [32]. Alternatively, a cine acquisition can be selected, which uses a higher radiation dose for a defined recording. This mode is applied when the catheter has been correctly positioned, after which the narrowing or blockage is captured. Another possible method for recording the medical condition is using the Digital Subtraction Angiography (DSA) mode. In this mode, an initial angiography is made, after which the contrast agent is injected. The initial image is then subtracted from subsequent angiographs, which removes any content that appears in both images, such as surrounding tissue. This allows for a clear visualisation of only the vessels that have been injected with the contrast medium, in order to accurately locate medical impairments [39]. Finally, a single-shot acquisition can be selected, which produces a still image. This technique is mostly used for determining values for parameters such as object thickness [32]. To initiate any acquisition, the physician must press a pedal underneath the surgical table, as can be seen in Figure 3.1. For moving images, the duration of the acquisition is equivalent to the amount of time the pedal is pressed. In order to record the corresponding data, no further action is required. When a suitable recording mode has been selected, recordings are saved automatically.

The used dataset contains data from all procedures that took place in the cath lab of the Reinier de Graaf hospital between the 28th of April 2017 and the 19th of April 2019. Within this time frame, the interventional X-ray system logged a total of 4,339 procedures. The procedures are registered by means of 137,843 acquisitions, where the number of acquisitions per procedure varies. Each acquisition is represented by 125 stored values, which correspond to the parameters logged by the interventional X-ray system. In a machine learning context, such parameters are referred to as features. The total dimension of the table that makes up the entire dataset is 137,843 by 125, which corresponds to the number of acquisitions and features respectively.

The program used to import and analyse the dataset is Matlab. When importing the Excel file that contains the interventional X-ray data, all but three features are ready for immediate use. The features that specify procedure date, start time and end time of the acquisition must first be transformed in order to be compatible with Matlab functions. After importing and transforming all acquisitions into Matlab, the next step is to filter the dataset as to exclude any procedures that are not CAGs. As this step is less straight-forward than initially anticipated, a sub-analysis is done to investigate possible methods to recognise CAG data within the dataset. This process is further explained in Section 3.1.2.

3.1.2. Procedure Type Determination

The dataset retrieved from the interventional X-ray system consists of all procedures that took place within the aforementioned period, including non-CAGs. Since the model in this report specifically aims to analyse the progress of a CAG, data from other procedures must be removed. Unfortunately, the dataset does not specify the procedure type, meaning that this specification must be extracted using a different method. Two applicable methods include the use of DoseWise data, or the extraction of information regarding procedure type from other features logged by the machine. The usability of each method is investigated and assessed below.

DoseWise Dataset

The Philips DoseWise Portal is an additional system that keeps track of each procedure and especially focuses on monitoring patient radiation dose. Unlike the interventional X-ray dataset, which is meant for servicing purposes such as maintenance and repair, DoseWise is a user-friendly program that continuously provides comprehensible insights throughout the procedure. It provides the control room with real-time information on patient radiation events [41]. Overall, the information logged by the DoseWise Portal is far less detailed than the information logged by the interventional X-ray system. Fortunately, it states the type of procedure that is being executed and is therefore suitable for validating procedure type within the interventional X-ray data. One entry from the DoseWise software represents one procedure and individual acquisitions are not included. Nevertheless, procedure type can be deduced by matching the dates and times between the two datasets. The acquired DoseWise dataset consists of 1,091 logged procedures that took place between the 5th of January 2018 to 4th of April 2019. When matching the dates and times of all procedures that specify the type to be 'CAG radialis' to the interventional X-ray dataset, a total of 601 overlapping procedures are found. The 601 procedures contain 16,536 acquisitions, which is more than adequate for the creation of a representative model using a random forest algorithm, given the number of features [3, 56]. Throughout the remainder of this report, only acquisitions from confirmed CAG cases are used.

It is possible that some of the logged procedures in the DoseWise dataset contain unanticipated CAG femoralis procedures, that were unregistered by DoseWise. According to cardiologist Dr. Hoftijzer, catheter insertion through the groin rather than through the wrist is exceptional and occurs in approximately 5% of the cases. However, validating whether this occurred is considered out of scope for this report. Deduction based on values of the logged features would not be reliable enough. Alternatively, individual patient records would have to be inspected. Nevertheless, the presence of unconfirmed CAG femoralis procedures within the dataset does not harm the model. In fact, it is desirable to implement these procedures rather than to exclude them from the dataset, in order to increase model suitability for realistic application where a diversion to the groin might also occur.

Application and Procedure Name

A possible disadvantage of using DoseWise data is that it might not be available for real-time implementation into the model. This means that analyses could only be performed retrospectively, whereas a desired application of the model is for intraoperative prediction of total procedure duration. An alternative method could be to indirectly deduce it from the features logged in the interventional X-ray dataset. Despite the fact that the data does not state the name of the procedure, several features provide information that suggest a correlation with the procedure type. The two features of the dataset that suggest promising abilities for determining procedure type include 'application name' and 'procedure name', as they specify techniques used and refer to certain internal and external body parts. In order to analyse this, the prevalence of each application and procedure name among CAGs and non-CAGs are presented in Table 3.1 and Table 3.2.

Table 3.1: The selected application names for CAGs and non-CAGs

Application name	Occurrences within CAG procedures	Occurrences within non-CAG procedures
Abdomen 3D	-	1
Abdomen ClarityIQ	19	13,100
Abdomen Thorax XperCT	-	358
Cardiac ClarityIQ Xpertaste	16,250	44,900
Cardiac EP	1	9,760
Cardiac Swing ClarityIQ	5	53
Head 3D	-	4
Head ClarityIQ	23	64
Head XperCT	-	1
Peripheral ClarityIQ	190	48,320
Thorax 3D	-	5
Thorax ClarityIQ	44	4,180

Table 3.2: The selected procedure names for CAGs and non-CAGs

Procedure name	Occurrences within CAG procedures	Occurrences within non-CAG procedures
1 Lower Leg 1 fps	14	10,410
1 Upper Leg 2 fps	93	18,580
2 Lower Legs 1 fps	7	1,385
2 Upper Legs 2 fps	7	1,495
Abdomen 2 fps	-	330
Abdomen 2 fps 25%	19	12,900
Abdomen 3 fps 25%	-	59
Abdomen 6 fps 25%	-	16
Abdomen Prop Scan	-	1
Abdomen SingleShot	-	132
Aortic Arch LAO 2 fps	23	-
Arm 2 fps	11	2,291
Cerebral 2 fps	-	67
Cerebral Prop Scan	-	1
Copy of Aortic Arch LAO 2 fps	-	4
Exposure 15 fps	-	7
Exposure 3.75 fps	1	9,340
Exposure 7.5 fps	-	413
Hand Foot 1 fps	-	907
Left Coro 15 fps Boost Xpertaste	1,304	4,252
Left Coro 15 fps Low	10,190	22,820
Left Coro 15 fps Normal Xpertaste	4750	17,630
Left Ventricle 15 fps	13	237
Pelvis/Iliac 2 fps	58	12,260
Swing LCA CRA 35 5.8s Xpertaste	3	1
Swing LCA CRA 30 5.3s Xpertaste	-	7
Swing RCA AP 4.0s Xpertaste	-	6
Thorax 2 fps	44	4,147
Thorax 6 fps	-	24
Thorax SingleShot	-	10
Xper CT Abdomen Roll (nurse)	-	28
Xper CT Cerebral LD	-	1

As shown in Table 3.1, the 'Cardiac ClarityIQ Xpertaste' application seems to be frequently used in both CAGs and non-CAGs. Therefore, the fact that this application name is selected does not mean that the procedure concerns a CAG. Since no other application name is commonly used in CAGs within this specific dataset, the type of procedure cannot be deduced from the feature that states application name. Regarding procedure name, three names are particularly common for CAGs, namely 'Left Coro 15 fps Boost Xpertaste', 'Left Coro 15 fps Low' and 'Left Coro 15 fps Normal Xpertaste'. Interesting is that all these names refer to the LCA, but since a CAG captures both arteries, it can be assumed that this setting is also used for the recording the RCA. Unfortunately, it appears that the aforementioned procedure names are regularly selected for non-CAG procedures as well. Therefore, the procedure name is also not a suitable indicator of procedure type.

Further analysis shows that certain combinations of application and procedure names are not appropriate indicators of procedure type either. In 188 procedures, more than one application name was selected. In most of these procedures, two different application names were used and in seven procedures, three application names were selected. Nevertheless, none of these occasions concerned a CAG. No combinations of application or procedure names are suitable to characterise the required procedure type. In conclusion, neither feature can be used to deduce procedure type and further investigation would be required to assess whether this can be done by means of other features or methods. Considering that the DoseWise software is capable of providing the necessary information, methods to determine procedure type have not been analysed further within this report. Instead, focus is put on the usability of interventional X-ray data for other purposes, such as predicting procedural workflow or total procedure duration. The next section describes the selected set of features used for the prediction model.

3.1.3. Machine Derived Features

Each acquisition in the dataset is represented by 125 different features. Some features represent general aspects of the acquisition, such as date and time. Other features concern technical aspects, such as radiation dose and other system settings. Positional features provide details on the exact position of the surgical table and the interventional X-ray system during each acquisition. A few features are related to personal characteristics, such as patient age. The complete set of features can be found in Appendix D. Out of the 125 features, 30 were selected to be relevant for predicting and analysing the workflow and progress of the procedure. The remaining 95 features were eliminated from the dataset because they were either incomplete or considered irrelevant for the aim of this particular project. The 30 selected features are presented and described in Table 3.3. For a better understanding of the configuration of the interventional X-ray system, Figure 3.1 presents its main components. This diagram should also help with understanding the features, as some of them refer to the particular components of the interventional X-ray system.

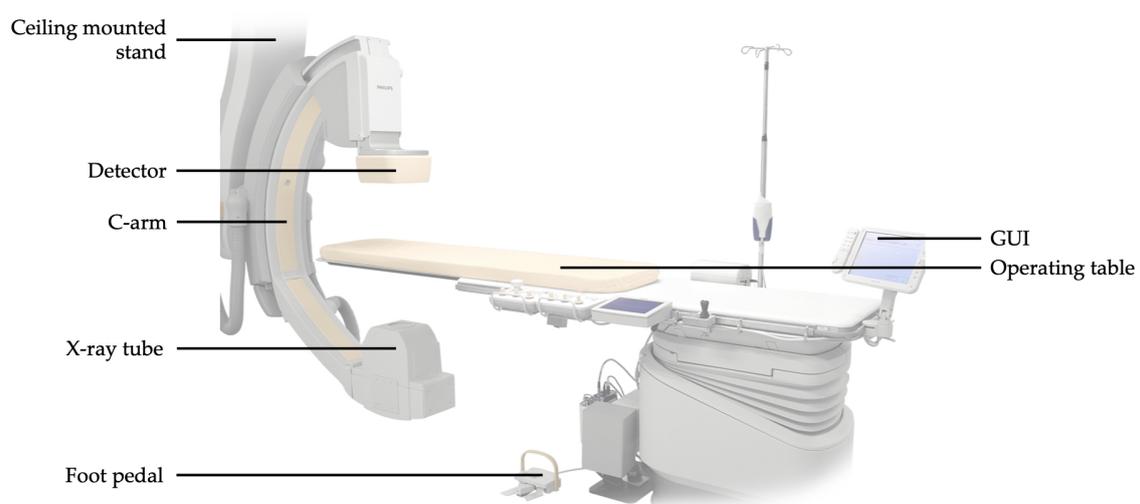


Figure 3.1: The main components of the interventional X-ray system. Original image retrieved from [40].

Table 3.3: The original features logged by the interventional X-ray system.

Code	Feature name	Unit	Feature explanation
OF1	Type	–	Acquisition type (fluoro, cine, DSA or Single-shot) [5, 24, 32, 48].
OF2	Application name	–	Type of application selected [37]. See Section 3.1.2.
OF3	Procedure name	–	Type of procedure selected [37]. See Section 3.1.2.
OF4	Run	<i>kV</i>	Voltage across the X-ray tube. One of the X-ray generator settings [37].
OF5	Run	<i>mA</i>	Emission current from the X-ray tube. The higher this number, the higher the number of photons reaching the detector, thus the higher the signal intensity.
OF6	Run	<i>ms</i>	Pulse width of the X-rays.
OF7	Acquisition time	<i>s</i>	Duration of the acquisition, referring to the total time during which X-rays were emitted by the tube.
OF8	Frame speed	<i>FPS</i>	Number of frames recorded by the image detector per second [32].
OF9	Object thickness	<i>cm</i>	Depth of the irradiated tissue. A larger bodypart requires a larger entrance dose to obtain the same detector dose [43]
OF10	Requested dose	<i>mGy</i>	Dose required by the detector to ensure sufficient image quality [43].
OF11	Dose ratio	%	Amount of X-ray dose measured by the image detector, with respect to the requested dose. [32, 43].
OF12	DAP	<i>mGy</i> <i>·cm²</i>	Dose Area Product. The dose of radiation delivered to the patient multiplied by the area of the exposed skin. Provides an estimated likelihood of skin damage.
OF13	Cumulative DAP	<i>mGy</i> <i>·cm²</i>	The sum of DAP values from all acquisitions, measured from the start of the procedure.
OF14	AirKerma	<i>mGy</i>	Measure of the amount of radiation energy deposited to a unit mass of air. Specifies the radiation concentration delivered to a point.
OF15	Cumulative Air-Kerma	<i>mGy</i>	The sum of AirKerma values from all acquisitions, measured from the start of the procedure.
OF16	Angulation start	°	The roll of the C-arm at the start of the acquisition. See Figure 3.2.
OF17	Angulation end	°	The roll of the C-arm at the end of the acquisition.
OF18	Rotation start	°	The propeller of the C-arm at the start of the acquisition. See Figure 3.2.
OF19	Rotation end	°	The propeller of the C-arm at the end of the acquisition.
OF20	SID	<i>m</i>	Source-to-image distance. The distance between the X-ray tube and the detector.
OF21	Position C-arm	° × 10 ²	The unfiltered angulation of the C-arm.
OF22	Position detector	° × 10 ²	The vertical position of the detector. Unfiltered version of the SID. See Figure 3.2.
OF23	Position propeller	<i>μm</i>	The unfiltered rotation of the C-arm.
OF24	Frontal beam long. position	<i>μm</i>	The longitudinal position of the X-ray system. See Figure 3.2.
OF25	Frontal rotation detector	<i>μm</i>	The position of the detector, moving between portrait and landscape position. See Figure 3.2.
OF26	Frontal z-rotation	°	The swing of the C-arm. See Figure 3.2.
OF27	Table height	<i>μm</i>	The vertical position of the operating table. See Figure 3.2.
OF28	Table lateral	<i>μm</i>	The lateral position of the operating table. See Figure 3.2.
OF29	Table longitudinal	<i>μm</i>	The longitudinal position of the operating table. See Figure 3.2.
OF30	Patient age range	–	The patient’s age group, defined by 10-year segments.

The positional features in Table 3.3 refer to specific directions that the interventional X-ray system and the operating table can move in. In order to clarify the implications of the stated directions, Figure 3.2 provides each positional feature code along with its specified direction.

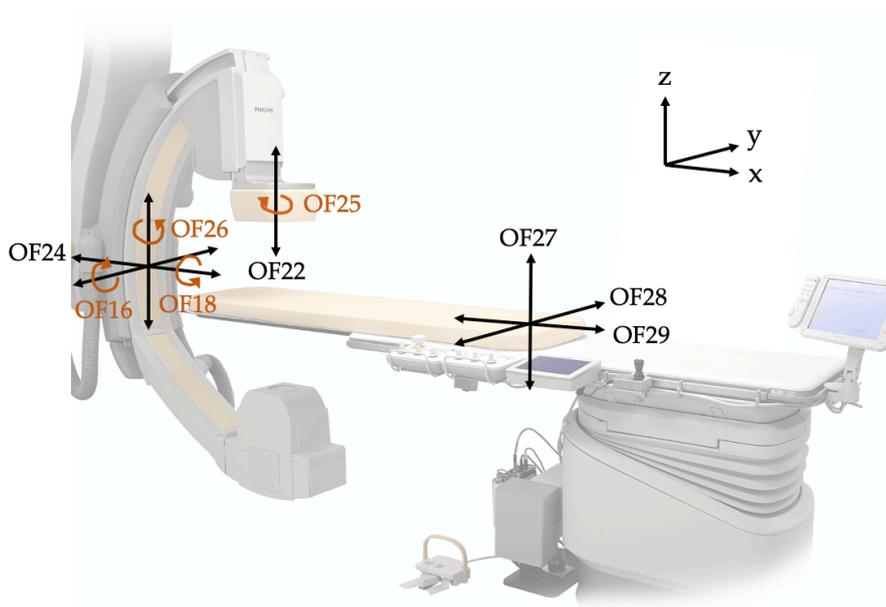


Figure 3.2: The directional axes of particular logged features. Original image retrieved from [40].

3.1.4. Additional Features

One of the disadvantages of using a random forest algorithm is that it does not take into account repeated measures. Repeated measures refer to the concept of repeating measurements of specific factors on the same subject, in this case the patient [55]. In the interventional X-ray dataset, multiple acquisitions are done on the same patient. However, random forest algorithms analyse individual data points, without considering prior acquisitions from the same patient. Information from preceding acquisitions could significantly benefit the analysis of the procedure. It is expected that progress estimations throughout the procedure of a particular patient will be more accurate when information regarding previous acquisitions is taken into account. In the original logged parameters, as presented in Table 3.3, no features provide any information regarding the procedure progress or previous acquisitions except for the two features 'cumulative AirKerma' and 'cumulative DAP'. In order to compensate for the lack of information from previous acquisitions, a total of 22 extra features have been generated and added to the input data. This ensures that the model takes into account the repeated measures that the random forest algorithm would normally disregard. In a later stage of the project, a feature importance analysis will show whether the additional features are indeed beneficial to the prediction model. The extra features and their definitions can be found in Table 3.4. Most of the extra features are calculated based on prior acquisitions of that particular procedure, using simple addition of cumulative values or occurrences relating to a specific feature. Furthermore, two features relate to the operative phase of the acquisition. The creation and implementation of these two features is explained in Section 3.1.5.

Table 3.4: The extra features added to the interventional X-ray dataset.

Code	Feature name	Unit	Feature explanation
EF1	Acquisition number	–	The total number of acquisitions since the start of the procedure.
EF2	Cumulative time	<i>s</i>	The elapsed time since the start of the procedure.
EF3	Time since acquisition	<i>s</i>	The elapsed time since the end of the previous acquisition.
EF4	Cumulative cine time	<i>s</i>	The total duration of all acquisitions for which the cine mode was used, within the procedure.
EF5	Cumulative fluoro time	<i>s</i>	The total duration of all acquisitions for which the fluoro mode was used, within the procedure.
EF6	Cumulative DSA time	<i>s</i>	The total duration of all acquisitions for which the DSA mode was used, within the procedure.
EF7	Number of cine uses	–	The number of acquisitions for which the cine mode was used, within the procedure.
EF8	Number of fluoro uses	–	The number of acquisitions for which the fluoro mode was used, within the procedure.
EF9	Number of DSA uses	–	The number of acquisitions for which the DSA mode was used, within the procedure.
EF10	Cumulative acquisition time	<i>s</i>	The total duration of all acquisitions, within the procedure.
EF11	Cumulative non-acquisition time	<i>s</i>	The total amount of time during which no acquisitions were made, within the procedure.
EF12	RCA phase	–	The estimated phase of the RCA recording. See Section 3.1.5 for more information.
EF13	LCA phase	–	The estimated phase the LCA recording. See Section 3.1.5 for more information.
EF14	Acquisition frequency	1/ <i>s</i>	The average number of acquisitions done per unit of time, within the procedure.
EF15	Vert. distance travelled table	μm	The total vertical distance travelled by the table since the start of the procedure.
EF16	Number of vert. moves table	–	The number of changes in the vertical position of the surgical table since the start of the procedure.
EF17	Lat. distance travelled table	μm	The total lateral distance travelled by the table since the start of the procedure.
EF18	Number of lat. moves table	–	The number of changes in the lateral position of the surgical table since the start of the procedure.
EF19	Long. distance travelled table	μm	The total longitudinal distance travelled by the table since the start of the procedure.
EF20	Number of long. moves table	–	The number of changes in the longitudinal position of the surgical table since the start of the procedure.
EF21	Long. distance travelled front beam	μm	The total longitudinal distance travelled by the frontal beam since the start of the procedure.
EF22	Number of long. moves front beam	–	The number of changes in the longitudinal position of the frontal beam since the start of the procedure.

3.1.5. Operative Phase Implementation

In prior work done for the overarching workflow project, K.M. van der Graaf defined the different phases of a CAG [28]. For each acquisition within the interventional X-ray dataset, the corresponding operative phase is estimated and used as additional input data for the model, with the aim of improving predictions of procedure duration. In Table 3.4, the operative phase is expressed by means of two features, EF12 and EF13, which refer to the phases for capturing the right and left coronary artery respectively. This subsection explains how the phases are detected within the interventional X-ray data and examines the detectability of each step. Furthermore, it contains an analysis of the overall accuracy and usability of the defined phases. In order to understand the numerical implementation of the operative phases into the additional feature set, it might be helpful to first attain deeper knowledge of the complete set of steps taken for a CAG.

1. Lab preparation

Start-up of the machines and readying of the cath lab for the first patient.

2. The procedure

Execution of the interventional procedure, including preparation and post-care of the patient.

(a) The patient preparation phase

Entry of the patient into the cath lab and preparation of the patient for the CAG.

(b) The operative phase

Execution of the CAG.

- i. *Realisation of endovascular access.*
- ii. *Catheter insertion and guidance to the upper aorta.*
- iii. *Entry and recording of the first coronary artery.*
- iv. *Entry and recording of the second coronary artery.*
- v. *Catheter removal and closure of the entry wound.*

(c) The post-care phase

Finalisation of procedure and exit of the patient from the cath lab.

3. The turnaround

Cleanup of resources from the previous procedure and preparation of the cath lab for the next patient.

4. The clean-up

Cleanup of the cath lab and the shutdown of machines [28].

K.M. van der Graaf defined steps 1 to 4 to be the high-level workflow phases. Additionally, steps 2a-c were presented as procedural phases. The second procedural step was further split up into steps 2b i-v, representing the operative phases. In addition to assigning qualitative definitions to each step, steps 2b iii-iv were given quantitative values. These substeps represent the recording of the RCA and the LCA, where recordings of the RCA generally require 3 steps, and capturing the LCA normally requires 5 sequential steps. According to K.M. van der Graaf, each of these steps correspond to a common range of angles at which the C-arm is positioned. Based on this observation, she expressed the steps in terms of angulation and rotation of the C-arm and hypothesised that the common values for angulation and rotation might be useful for detecting the operative phase of the procedure. By implementing features that state the detected steps of the LCA and RCA phase into this classification model, this hypothesis can be tested. The steps and corresponding values for angulation and rotation are shown in Table 3.5, under 'original angulation' and 'original rotation'. The intervals are expressed using terms that specifically relate to the spinning direction of the C-arm. For example, caudal angulation (CAU) refers to a tilt of the central ray, projected from the tube underneath the operating table towards the feet of the patient as shown in Figure 3.4. For cranial angulation (CRA), the central ray is tilted towards the opposite direction, pointing towards the head of the patient. At 0°, the central ray is emitted vertically upward. In terms of rotation, a Left Anterior Oblique (LAO) view points the central ray towards the left side of the patient. For the Right Anterior Oblique (RAO), the tube is tilted as to capture the right side of the subject, for which the detector moves to the right accordingly [36]. Both rotational directions can be seen in Figure 3.3.

Table 3.5: The general angulation and rotation values of each operational step [28].

Phase	Step	Original angulation	Applied angulation	Original rotation	Applied rotation
RCA	1	0°	[-10°, 10°]	[30°, 40°] LAO	[20°, 50°]
	2	[30°, 40°] CAU	[20°, 50°]	[30°, 40°] RAO	[-50°, -20°]
	3	[30°, 40°] CRA	[-50°, -20°]	[30°, 40°] LAO	[20°, 50°]
LCA	1	0°	[-10°, 10°]	[30°, 40°] LAO	[20°, 50°]
	2	[30°, 40°] CAU	[20°, 50°]	[30°, 40°] LAO	[20°, 50°]
	3	30° CRA	[-40°, -20°]	30° LAO	[20°, 40°]
	4	[30°, 40°] CRA	[-50°, -20°]	[0°, 20°] RAO	[-30°, 10°]
	5a	20° CAU	[10°, 50°]	20° RAO	[-30°, -10°]
	5b	40° CAU	[10°, 50°]	20° RAO	[-30°, -10°]

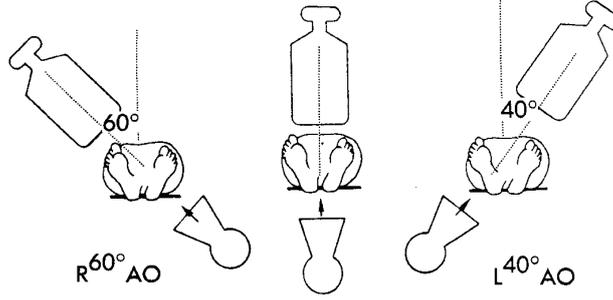


Figure 3.3: The right and left anterior oblique rotation [51].

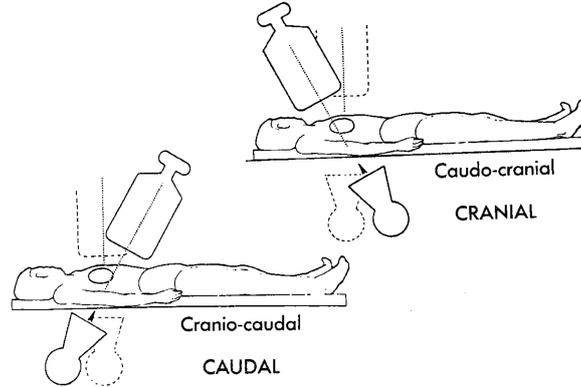


Figure 3.4: The caudal and cranial angulation [51].

For each acquisition, the RCA feature (EF12) states which combinations of angulation and rotation from the RCA steps in Table 3.5 have been detected within the entire procedure so far. The same goes for the LCA feature (EF13), based on the corresponding values for angulation and rotation from the LCA steps. The standard value of both features is 0. As soon as the values for angulation and rotation of the acquisition match the values from the first step of either the RCA or LCA phase, the value of the corresponding feature changes to 1. Then, the model scans for a match between the subsequent acquisitions and values for angulation and rotation of the second step. The model is configured to require the detection of one step before being able to scan for a match with the next step. One feature tests for indications that suggest a recording of the RCA, whereas the other feature aims to detect LCA acquisitions. Most physicians first examine the RCA, followed by the LCA [28]. However, as this sequence of events cannot be guaranteed, the analysis must test each acquisition for both phases. Therefore, the two operative phases are simultaneously included in the interventional X-ray dataset.

In order to increase detectability, the angulation range and rotation range are given an extra margin of $\pm 10^\circ$. This implies that the value of the RCA feature switches to 1 if the angulation values are within 10° CRA and 10° CAU and the rotation values are within LAO 20° and 50° . The same goes for the other RCA and LCA phases and their respective angulation and rotation values. All intervals that correspond to the extended margins are shown under 'applied angulation' and 'applied rotation' in Table 3.5. Furthermore, the interventional X-ray system does not record the angles using the terms shown in Figure 3.3 and Figure 3.4, but uses negative values to express the CRA and RAO directions instead. The intervals that include the extended margins are expressed using the format that corresponds to the interventional X-ray data. At last, step 5b from Table 3.5 is applied when 5a does not provide an adequate recording. For simplicity purposes, the substeps are merged, corresponding to an angular range of 10° to 50° and rotational range of 10° to 30° .

Detectability of the operative phases

The histograms in Figure 3.5 show the distribution of detected phases for all acquisition within the dataset, according to the intervals from Table 3.5. Based on the used dataset, there is no method to retrieve the true phase of each acquisition. Nevertheless, the extent to which each of the defined phases can be detected and might contribute to the predictions for procedure duration can be analysed. It appears that no steps from either phase are detected for approximately 2,200 of the acquisitions. A possible clarification is that these instances correspond to acquisitions in which the other coronary artery was being examined. A different option is that they are acquisitions during which the catheter had just arrived in the aorta arch. When this is the case, the interventional X-ray system is usually in rest position and both angulation and rotation values correspond to 0° . Alternatively, the angulation and rotation values are different from the common values for other reasons. Such reasons can merely be speculated and validating this was considered to be beyond the scope of this research.

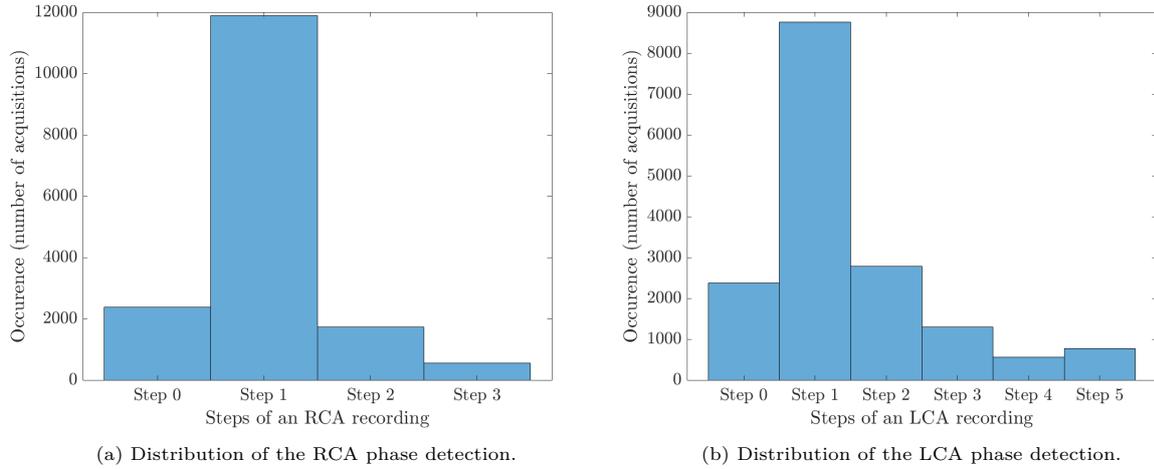


Figure 3.5: The detection rate of each coronary artery phase within the dataset.

Step 1 is by far the most detected step in the dataset. It should be noted that for both the RCA and the LCA phases, the first step corresponds to the same range of values for angulation and rotation. Therefore, any step that has been detected as being step 1 for the RCA, has also been classified as step 1 for the LCA. More information would be required to validate whether the step concerns the RCA or the LCA. Furthermore, the extra margins of 10° added to each step result in more overlap between intervals that correspond to each step. Therefore, steps could be confused with one another. However, without the use of the aforementioned margins, only a minimal number of phases have shown to be detected using the available data. This shows a variability in angulation and rotation values, beyond the defined range of values given in Table 3.5.

To be able to analyse the detection of each step with respect to the procedural progress, progress is split into segments of 10%. For each segment, the steps of all acquisitions within that segment are given. This provides an insight into the moment at which specific steps are detected throughout the interventional procedure. Figures 3.6 and 3.7 show that in the first 10% of the procedure, the model mostly detects either no phase or the first step of the phases. Both RCA and LCA phases show the same increase in detection rate of the first phase between 0% and 40% of the procedural progress, as expected since step 1 is the same for both phases. Furthermore, Figure 3.6 shows that the detection rate of the first step of the RCA phase also increases between 70% and 100% of the interventional progress. Given the fact that the angular and rotational range is the same for step 1 of both phases, one would expect the same rise in the graph showing the detected LCA steps. Nevertheless, this is not the case. The reason might be that subsequent acquisitions are found to correspond to higher steps of the LCA phase. The fact that steps from the LCA phase lie closer together in terms of angular and rotational values could enable this. This means that steps that are identified as step 1 are more likely to fit the angular and rotational range of step 2 or higher. Alternatively, the defined range of values for steps of the LCA recording are more accurate with respect to the true values for the position of the C-arm. Both reasons could explain why the LCA phase shows a high detection rate for steps beyond the first step, whereas the majority of the detected RCA steps correspond to step 1, even at the end of the procedure. This suggests a low detectability of the second step of the RCA phase. The fact that there are generally fewer acquisitions between 30% and 70% of a procedure, as shown in Appendix A, could also contribute to the fall and subsequent rise in detection rate within this procedural time frame.

The fact that step 5 from the LCA phase and step 2 from the RCA phase are relatively similar in terms of angulation and rotation could explain why the latter shows an exponential increase beyond 90% of the procedural progress. Following this increase in detection rate of step 2 from the RCA phase, the model now seeks to detect step 3, resulting in an increased detection rate of the third step as well. The LCA phase shows a more gradual increase in detection rate of all subsequent steps, with respect to procedural progress. This observation further supports the theory of the low detectability of step 2 of the RCA phase. This would imply that the angular and rotational intervals, as shown in Table 3.5, are presumably not representative of the position of the C-arm in the RCA phase. If this assumption were correct and the detected LCA steps are indeed more reliable, it is possible that the increased detection rate of the LCA steps towards the end of the procedure is due to the fact that the LCA recordings are usually performed after imaging of the RCA.

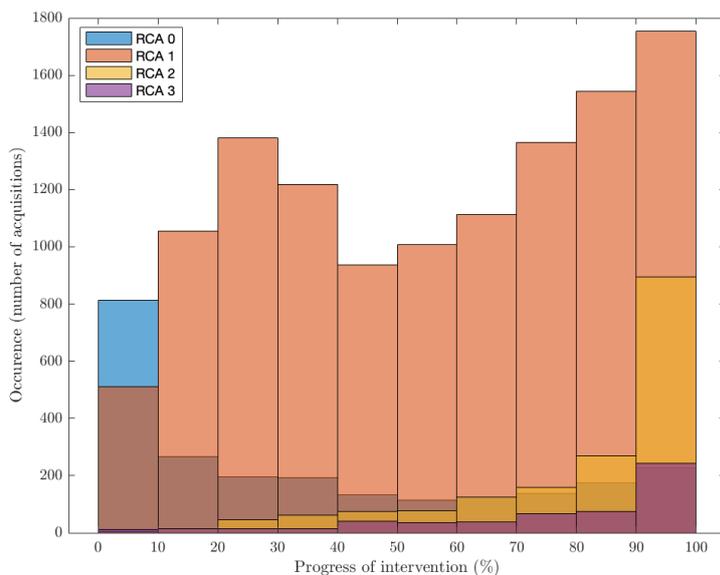


Figure 3.6: The detection rate of each RCA phase with respect to progress.

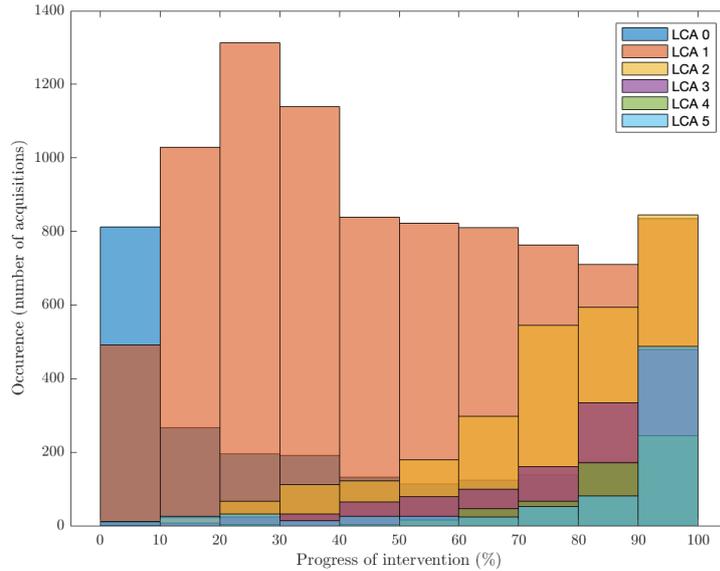


Figure 3.7: The detection rate of each LCA phase with respect to progress.

Usability of the operative phases

In conclusion, the observations do not show enough evidence to state that the angulation and rotation are beneficial for estimating the progress of the interventional procedure. The steps of the LCA phase show a higher detectability, yet the detection accuracy cannot be guaranteed using the information that is currently available. For an accurate analysis, the steps must be represented by more features, as the angulation and rotation are too variable to be detected by the original intervals presented in Table 3.5. When adding margins, the intervals are too similar and cannot be differentiated without extra information. Even though the success of the features that detect the RCA and LCA phases cannot be guaranteed based on this analysis, the features are still implemented into the random forest classification model. If the model finds the additional features to be invaluable, it will automatically remove them from the feature set during the recursive feature elimination procedure. This is where the optimal number of features to be implemented into the model is deduced, as well as the exact list of features that should make up the final dataset. The steps for optimisation of the feature set are further elaborated upon in Section 3.3.

3.1.6. Data Framing

In order to move on to the creation of the classification model, the input dataset must be tested for noteworthy outliers and split into training data and test data. The reason for these steps is explained below.

Outlier removal

Two procedures within the dataset contain acquisitions that are separated by a significant amount of time. One of these procedures contains multiple acquisitions within the first 12 minutes of the procedure, followed by one last acquisition executed 58 minutes later. The other procedure shows a similar situation. Since the procedures show only one acquisition after a relatively large amount of time, it is likely that the datapoint is a mistake and is not part of the same procedure. This assumption has been acknowledged by a cardiologist at the Reinier de Graaf hospital. It is possible that the new patient file had not yet been uploaded. Alternatively, the foot pedal could have been pressed for cleaning purposes, rather than to capture an acquisition. To ensure that the results would not bias the training and outcome of the classification model, all acquisitions from the two procedures that possibly contained faulty data are removed from the dataset.

Further outliers in the data are a result of delayed procedures. Nevertheless, these procedures do not show a questionable amount of time in between different acquisitions and are therefore assumed not to be a mistake in the data. Furthermore, these outliers must be included within the data used to train the classification model, in order to enable the detection of delayed procedures in the future. This trains the model to understand how different features change when the procedures show a potential delay. Additionally, random forest is said to be robust to bias due to outliers, due to its use of bagging. Seeing that each tree is trained using a different subset of data, the outliers are only included in the training data for a number of trees. This reduces the extent to which the model is biased towards the outliers [3, 4].

Data splitting

In this report, the random forest model is optimised and tested by means of two different methods. This is done because both methods provide interesting and sometimes varying insight into classification accuracy and overfitting. The first method evaluates the out-of-bag error to assess classification accuracy. The advantage of out-of-bag analysis is that all trees are tested using different subsets of data, which means that the resulting accuracy is based on a larger variety of data. For example, if the test data from the cross-validation method is not representative of the overall dataset and contains outliers that the training data does not contain, the classification accuracy will be biased. When out-of-bag analysis is applied, the presence of outliers in the test data has a lower effect on the classification accuracy, since approximately one third of the trees are not tested using data that contains the outlier [3]. Furthermore, out-of-bag analysis removes the need to split up the dataset into training and test data. This implies that all data can be used to train the model, increasing its predictive abilities. This is also why the out-of-bag method is particularly suitable for small datasets. Additionally, out-of-bag analysis has a relatively fast computational speed. The tree can be built and evaluated simultaneously. Nevertheless, each datapoint from the test data is only tested on the trees that did not use the datapoint to be constructed. This means that each datapoint is evaluated using a smaller number of trees compared to the second validation method, which could lead to a lower classification accuracy [25].

The second method is based on cross-validation using test data. This method implies that once the model is completed and optimised, results are tested using data that was not used to train the classification model. Therefore, each datapoint from the test data can be evaluated using every decision trees within the random forest. Another advantage of cross-validation is that the method allows for comparing the training error to the test error. When the model outputs a training error that is close to zero and classification error based on test data is significantly higher, the model is too specifically trained for the training data and overfitting has occurred. This implies that an excessive amount of the noisy data was taken into account when creating the classification model. The random forest model now requires generalisation to make it more suitable for unseen test data. This can be done by changing the hyperparameters, as further explained in Section 3.2.3. Furthermore, cross-validation is suitable for the testing the model using completely new datasets. For example, cross-validation allows for testing of the classification model designed in this report using interventional X-ray data from other time periods, without having to change the classification model itself.

In this report, 15% of the datapoints are removed from the original dataset to be used as testing data. Training points are randomly selected without replacement, to ensure that none of them are included in the test data. For the remainder of this chapter, only the training data is used as input data for the creation and optimisation of the random forest model. This means that both validation methods are applied to the same model, created by using the training data. This enables evaluation of an identical model, despite the validation method used. Only for the distribution analysis of procedure duration and the selection of classes, the entire dataset is used. These steps are further elaborated on in Section 3.2.1 and Section 3.2.2 respectively. Once the model has been tested using both validation methods, the validation method that results in the highest classification error is used to evaluate the model in further depth. This is done to reduce the optimistic bias of the results to the fullest extent.

3.2. Classification Test Setup

This section revolves around the preparation and optimisation of the random forest classification model. In order to predict the procedure duration based on classification, the durational range of each class must be defined in advance. The distribution of procedure duration within the dataset is analysed in Section 3.2.1, based on which the classes are selected in Section 3.2.2. Finally, Section 3.2.3 explains the optimisation of the random forest model, by means of hyperparameter tuning.

3.2.1. Sample Distribution

The aim of the classification model is to predict the total procedure duration for each individual acquisition. The procedure duration refers to the time between the first acquisition and the last acquisition, as this is the time frame logged by the interventional X-ray system. The other phases of the procedure, including patient preparation and the post-care phase, are not considered in this particular model. Nevertheless, most variation occurs in the operative phase, thus a large part of the indicators of a deviating procedure duration should be detectable by the model [28]. Since the model is used to classify the estimated duration of each procedure, a number of preset classes must be defined, which is done based on the distribution of data. In order to analyse this distribution, all procedures included in the available dataset are arranged according to their value for total procedure duration. In order to account for the fact that the procedures contain different numbers of acquisitions, one data point per procedure is considered for the analysis. The distribution of procedure duration can be seen in Figure 3.8.

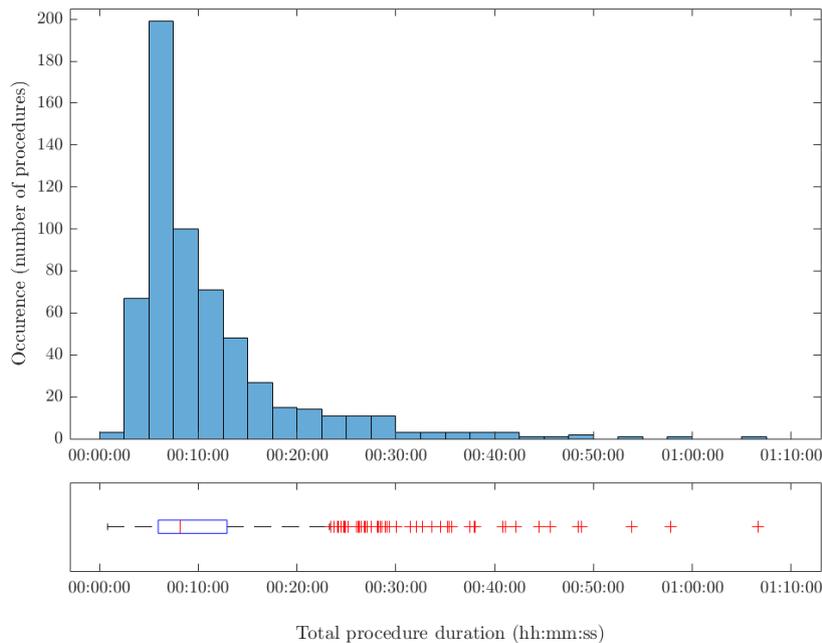


Figure 3.8: The distribution of procedure duration within the dataset.

In this particular histogram, each bin width represents a range of 2.5 minutes. For the majority of the procedures, less than ten minutes were required to complete the examination of both arteries. A large part of the procedures require only 5 to 7.5 minutes and a few examinations were even executed within 2.5 minutes. Nevertheless, several cases appear to have lasted longer than an hour, which makes it clear that the total duration varies significantly. Underneath the histogram, a boxplot is used to clarify the distribution and presence of outliers. The blue box represents the interquartile range, referring to the middle 50% of data. All values outside 1.5 times the interquartile range are considered outliers, corresponding to a procedure duration longer than 23 minutes. The median value for procedure duration is 8 minutes and 9 seconds and the mean procedure duration is exactly 11 minutes.

3.2.2. Class Selection

Appropriate procedure duration classes can be selected based on the distribution in Figure 3.8. The aim is to select four classes that represent procedure duration and correspond to a standard duration, slight delay, significant delay and severe delay. A possible option for defining the four classes is by using the mean duration as a reference point and adding one, two and three standard deviations respectively. This would mean that the classes are represented by the spaces within and outside the purple, green and blue lines on Figure 3.9. Nevertheless, the mean is not an appropriate reference point because it is skewed towards the outliers and is relatively far off from the most common value for procedure duration. Using the median value to represent the standard duration of a procedure is more suitable in this case. Furthermore, the approximate value of the first standard deviation from the mean is 20 minutes, but the vast majority of the procedures requires less time to complete the examination. Therefore, extra separation is required for more accurate predictions within this time frame. Given that the median value for duration is approximately 8 minutes, the first two classes are selected to represent the range between 0 to 10 minutes and 10 to 20 minutes. An appropriate duration range for the third and fourth class correspond to a duration of 20 to 30 minutes and a duration greater than 30 minutes respectively.

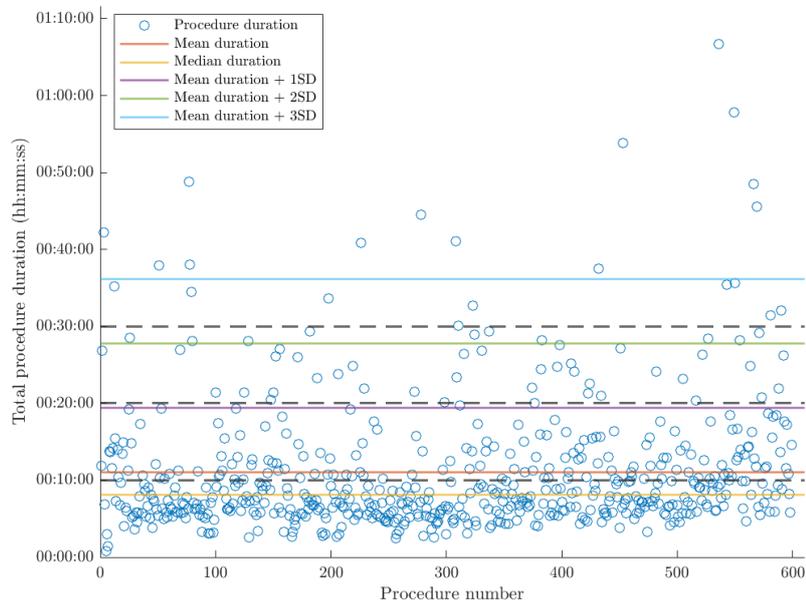


Figure 3.9: The distribution of procedure duration with respect to individual procedures and selected classes.

A distribution of all 599 procedures and their respective duration can be seen in Figure 3.9, along with the selected classes separated by the dashed lines. From the figure above, it is clear that the standard class shows the highest density in terms of procedures. Nevertheless, the other classes also contain a significant number of procedures, which is required for adequate training of the random forest classification model. The number of procedures per class is examined and presented in Table 3.6. The table shows that Class 1 is by far the most prevalent class, even though it corresponds to a duration range that lies below the mean.

Table 3.6: The selected classes for procedure duration and occurrence within the dataset.

Class	Duration (minutes)	Number of procedures
1	0-10	369 (61.6%)
2	10-20	161 (26.9%)
3	20-30	47 (7.85%)
4	>30	22 (3.67%)

Even though Class 3 and Class 4 are represented by a relatively small number of cases, they notably contribute to the random forest model. The use of random forest is particularly appropriate for small sample sizes, due to its application of bootstrap aggregation. By randomly selecting a subset of Class 3 or 4 data with replacement, there are no restrictions on the size of the training dataset, while maintaining the characteristics of the original dataset. This ensures that an appropriately sized representation of the data can be implemented into the model. Nevertheless, use of a smaller training dataset does reduce model performance with respect to a bigger set of training data. In this case, the selection of classes is based on the preference to optimise the model for detecting the severity of the delay, rather than small deviations in terms of a few minutes. If the latter were the case, smaller margins would have been selected, closer to the start of the procedure. As this is the first step of the analysis of this particular dataset, broader margins seem more appropriate. Nevertheless, using smaller classes could be an interesting modification in follow-up studies.

3.2.3. Hyperparameter Tuning

In order to set up the classification model, multiple parameters that characterise the random forest must be defined. Such parameters are referred to as hyperparameters, whose values depend on the input data and the aim of the model. Optimisation of the hyperparameters can maximise prediction accuracy. However, incorrect hyperparameter tuning can lead to overfitting. This section defines the concept of overfitting and explains how it can be prevented, after which each hyperparameter is described and optimised.

Overfitting

Generalisability is one of the main objectives of machine learning [10]. Overfitting occurs when the model is too specifically trained for the training data and performs poorly when exposed to new data, which implies a low generalisability. When the noisy data and outliers are overly implemented into the classification model, it becomes biased and shows a reduced performance on unseen data. The occurrence of overfitting can be tested by running the model on the training data, followed by the test data. If the training error is close to zero and the test error is significantly larger, the model is probably overfitted. For out-of-bag evaluation, overfitting can be shown by means of a high classification error. However, the error could in this case also be due to other reasons. In general, random forest techniques have a low probability of overfitting, with respect to other machine learning methods [4]. Since the use of bagging reduces the influence of noisy data and outliers, as explained in Section 3.1.6, the classification model is less likely to be susceptible to dominant data. This results in a generalised model that is suitable for use by new data.

In order to reduce the likelihood of overfitting to a minimum, the classification model should be defined by the use of optimal hyperparameters. Ideally, the model is detailed enough to provide an accurate prediction of procedure duration. Nevertheless, it should also perform well when tested on new data, in order to allow for predictions of future procedures. Hyperparameters determine the behaviour of the algorithm and affect its randomness and performance. Standard programmes are known to provide an adequate set of default hyperparameters [14, 16]. Nevertheless, the application of an algorithmic optimisation could improve the selection of hyperparameters even further. For each of the hyperparameters discussed in this section, multiple options are tested with respect to their out-of-bag error. The out-of-bag error is used due to the fact that it is computationally efficient and can be tested simultaneously to model creation. For each parameter, the value that corresponds to the lowest out-of-bag error should be used. The error reduces as classification accuracy increases, yet the error will go up again as overfitting starts to occur. Graphical demonstrations of this can be seen in the next few sections.

Number of trees

The first hyperparameter involves the number of decision trees used in the random forest model. As more individual trees are built and trained by the model, more subsets of the training data are used to optimise predictive ability, resulting in a higher classification accuracy. Therefore, the number of trees must be sufficiently large, in order to prevent high error rates. Nevertheless, using more decision trees leads to a significantly higher computation time, which increases linearly with respect to the number of trees. As a rule of thumb, ten times the number of features used in the model can be used as a starting value for the number of trees [9, 16]. At this point, the number of features implemented is 52.

The sufficient number of trees required for the classification model can be deduced by graphing the number of trees against the out-of-bag misclassification probability, which refers to the probability that a data point is incorrectly classified with respect to its true duration class. When the misclassification probability stabilises, the appropriate number of trees has been reached. When running the data through a simple random forest classification model and setting the maximum number of trees to 600, as to ensure that more than ten times the number of features are used, Figure 3.10 is produced.

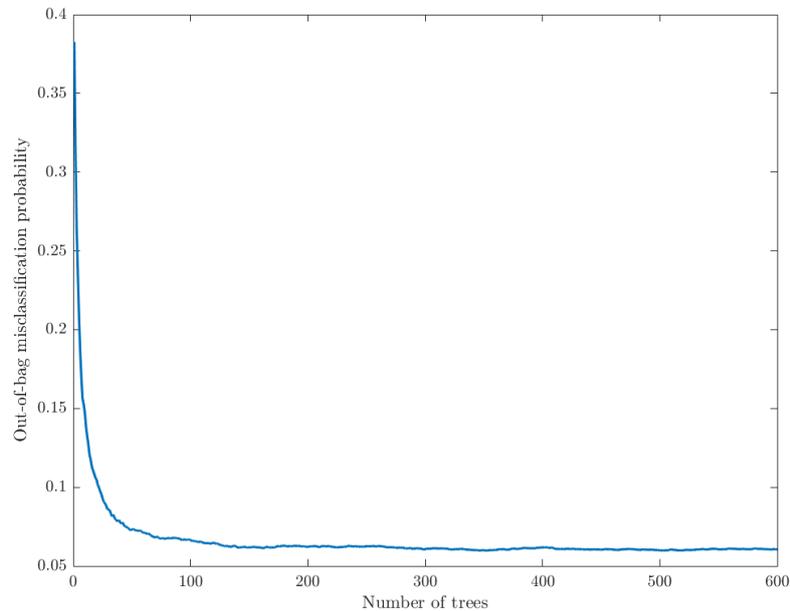


Figure 3.10: The effect of the number of trees on the out-of-bag classification accuracy.

From this graph, it can be seen that from the point when approximately 150 decision trees are used, the error rate stabilises. However, error still slightly reduces as the number of trees increases, especially up to 300 trees. Even though the error rate never stabilises completely, deviations are minimal and do not affect the misclassification probability enough to outweigh the increase in computation time. A total of 300 decision trees are used for the remaining optimisation of the random forest classification model.

In-bag fraction

The second hyperparameter is the in-bag fraction, which refers to the size of the sample set of observations selected during the bootstrap aggregation process. Using a bigger sample set means that more data is used to train the model, which will probably result in a higher classification accuracy. Nevertheless, a higher in-bag fraction reduces the randomness of the model, which could lead to overfitting. The default value for in-bag fraction is 1, which means that the entire dataset of observations is used. Nevertheless, as replacement is used, the bootstrap sample is different from the training set. The observations are selected with replacement, meaning that some observations are selected multiple times whereas others do not contribute to the sample whatsoever. The latter observations are the so-called out-of-bag observations. On average, two-thirds of the original dataset end up in each bootstrap sample, when replacement is used [3]. In order to test for the optimal in-bag fraction (IBF) for this particular dataset and selected number of trees, four different fractions are tested with respect to the out-of-bag misclassification probability, as shown in Figure 3.11. From 25 trees onwards, an in-bag fraction of 1 results in the lowest misclassification probability and is considered the most appropriate value. The occurrence of overfitting can be ruled out, as this would have increased the misclassification probability notably, as described in Section 3.1.6.

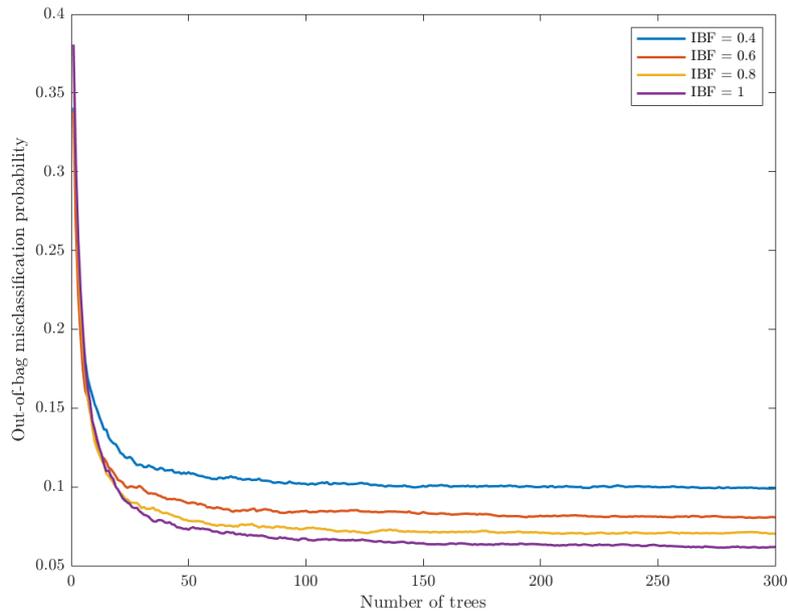


Figure 3.11: The effect of the in-bag fraction on the out-of-bag classification accuracy.

Number of features to sample

An essential characteristic of the random forest algorithm, as defined by developer L. Breiman, is that at each decision split only a subset of the features are considered for further splitting [4]. The size of this subset is defined by the number of features to sample [35]. This prevents dominant features from overshadowing other possibly relevant features and increases the stability of the model by reducing variance. This in turn reduces the chances of overfitting [18]. In this context, variance refers to the sensitivity of the classification model to specific characteristics of the dataset, such as noise and specific observations [6].

Variance can be decreased by increasing the number of features to use within a bootstrap sample, as to reduce randomness. However, using too many features in a bootstrap sample could lead to overfitting. In classification, the default value for the number of sampled features is the square root of the number of features. With the current feature set of 52 features, this means that eight features would be used for each decision tree. Different values for number of features to sample are tested against out-of-bag misclassification probability as shown in Figure 3.12. These values correspond to 80%, 60%, 40%, the square root and 10% of the number of features respectively. From this graph, it appears that when a small number of features is sampled, such as 5, the model is not accurate enough. Furthermore, when a high number of features is sampled, the misclassification probability increases and the model is probably subject to overfitting. The optimal number of features to sample on the current dataset and feature set, out of the tested values, is 21 and corresponds to 60% of the feature set size. Despite its promising results, using a set of 21 features at each decision split increases the computation time drastically. When running the random forest classification model on the entire dataset and considering a feature set of size 21 at each split, the model takes 148.7 more seconds to run. For the recursive feature elimination, this would result in approximately 21 extra hours of extra computation time. When using the default value for number of features to sample, computation time is reduced by 18%. When put into perspective, this leads to a decrease in out-of-bag misclassification probability of less than one percent. This change does not outweigh the significant decrease in computation time. Therefore, the default value, equal to the number of features squared, is used as the number of features to sample throughout the remainder of the optimisation model.

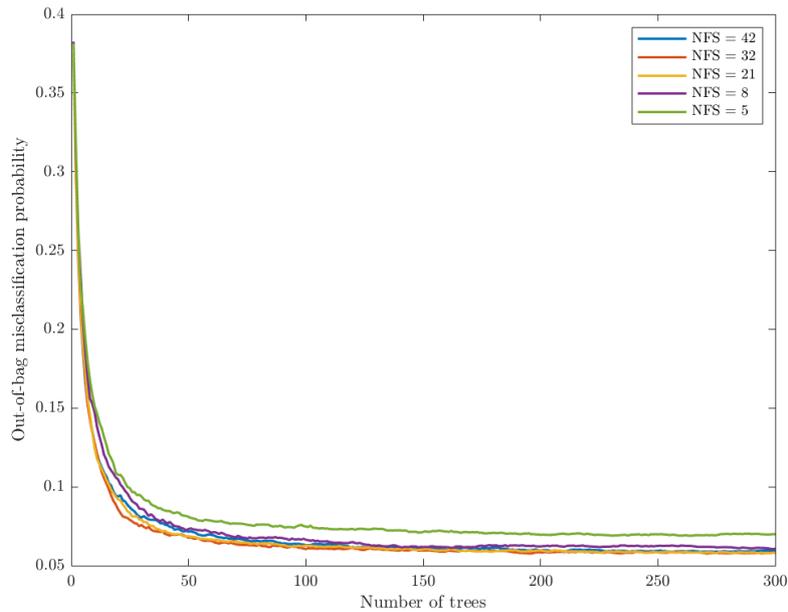


Figure 3.12: The effect of the number of sampled features on the out-of-bag classification accuracy.

Tree depth

The tree depth refers to the number of layers in each individual decision tree. Increasing the number of layers in a decision tree increases the accuracy of the model and ensures that all necessary features are taken into account. However, having too many layers could lead to overfitting. Furthermore, a high number of layers reduces interpretability of the model, although this is unlikely to be of importance when applying the decision trees to a bigger structure such as a random forest. In other programming languages such as Python or R, the tree depth can be adjusted manually. However, Matlab does not allow for this and controls the tree depth by means of three other parameters, including the maximum number of branch node splits, the minimum leaf size and the minimum parent size, as discussed below.

Maximum number of splits

The maximum number of splits determines the maximum variation of binary pathways that a given node can split into. In order to increase tree depth, a large number of possible splits must be used [21]. The default value is $n - 1$, where n is the length of the dataset. The total number of possible splits equals $2^{k-1} - 1$, where k equals the number of classes. As there are four possible duration classes, there are seven possible splits. In order to ensure all possible splits are included in the model, the default value of seven is maintained.

Minimum leaf size

The minimum leaf size is the minimum number of observations required per leaf node and determines, among other parameters, when to stop the splitting process. Having a small leaf size means that the tree will keep growing to split up the observations into further leaf nodes. For example, an input dataset that consists of 1000 data points is split up into 600 and 400 data points. If the minimum leaf size is set to 400, this condition has been satisfied and the tree can be considered finished, although probably with a poor predictive performance. However, if the minimum leaf size is 60, the tree can keep growing until 60 observations or more are left in the leaf node, which would probably result in a higher accuracy. [4]. By default, a classification tree requires one observation per leaf [35]. Besides a higher accuracy, a small leaf size may lead to overfitting, since noisy data is now also given individual leaf nodes. The optimal leaf size can be found by graphing multiple random forest estimations against the out-of-bag misclassification probability, using different leaf sizes.

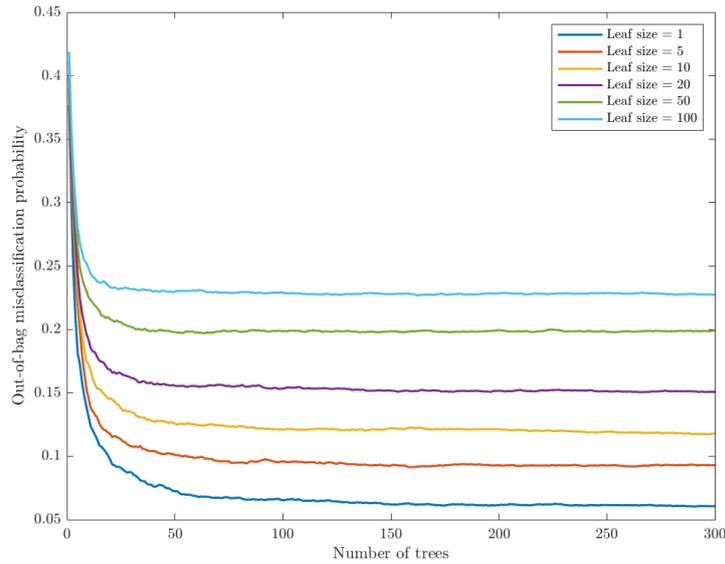


Figure 3.13: The effect of the minimum leaf size on out-of-bag classification accuracy.

Figure 3.13 shows that when a higher number of leaves is used, the model requires a lower number of trees to stabilise the error rate. However, a higher number of leaves results in a higher out-of-bag misclassification probability. The misclassification probability is lowest when the minimum leaf size is 1. The graph shows that when this leaf size is used, overfitting does not occur as the misclassification probability stays low. Therefore, a leaf size of 1 is used for the remaining model optimisation steps.

Minimum parent size

The minimum parent size defines the number of observations in each parent node before a split is attempted. This value must be small to get a large tree depth. Decreasing the minimum parent size increases the predictive abilities of the model, but could lead to overfitting. The default value of minimum parent size is twice the value for minimum leaf size, which means that a parent node would be required to have at least two observations before further splitting. As it has been shown that overfitting does not occur when the minimum leaf size is used and no restrictions are imposed on the minimum parent size, the default value is maintained.

3.3. Recursive Feature Elimination

Running a random forest model with too many features can lead to overfitting, resulting in a model that is too excessively trained on the training data and will perform poorly once exposed to new data. Furthermore, using too many correlated features impact the model’s ability to recognise strong predictors [11, 45]. In order to find the optimal number and combination of features for the random forest classification model, a recursive feature elimination (RFE) method is applied. Using this iteration method, a backward selection of important features is made by continuously removing the least important feature for determining duration class. Initially, all 52 features are used as input for the random classification model. The feature that has shown to be least significant for estimation of the duration class is then removed from the feature set, along with its respective data and the model is run again. For each new feature set, the misclassification error is evaluated and the removed feature is registered. This process is repeated until only one feature remains, after which the misclassification errors are graphed against the number of features used by the model. The lowest point of this curve represents the lowest misclassification probability and corresponds to the optimal number of features. In order to provide a more robust conclusion on the optimal number of features, as well as the optimal set of features, the RFE is run ten times. Furthermore, two different types of error are used. For the first five runs, the out-of-bag error is used to evaluate misclassification probability and eliminate features. For the other five, cross-validation is used. In all ten RFE runs, different randomly selected subsets of training data are used, as to avoid selection bias. The original training to test data ratio of 0.85 to 0.15 is maintained. The results of all ten RFE runs are plotted in Figure 3.14. [14].

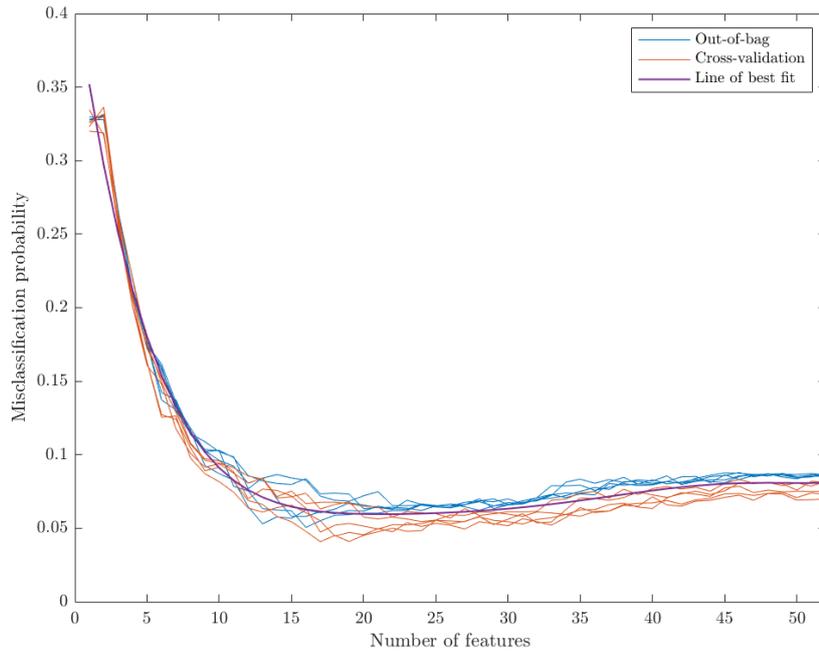


Figure 3.14: The misclassification probability against the number of features, based on 10 RFE applications.

The optimal number of features is deduced by taking the average of all ten optima, resulting in a set of 19 features. To find the optimal combination of features that should make up the final set, each feature is given a score based on its final ranking in each of the RFE applications. The mathematical expression for the feature score of feature i is given in Equation (3.1).

$$Score_i = \frac{1}{\sum_{n=1}^N Fi_n} \quad (3.1)$$

In Equation (3.1), F represents the ranking of feature i in one specific RFE application and N represents the total number of times the RFE is applied. The process is repeated for each feature in the feature set. The 19 features with the highest scores are displayed in Table 3.7. This is the final feature set for estimating procedure duration class using the random forest model. Out of the 19 most significant features, 13 correspond to features from the original dataset. The remaining features were added in Section 3.1.4. The complete result of the RFE can be found in Appendix C.

Table 3.7: The optimal feature set based on results from the RFE.

Code	Feature name	Unit	Score
EF2	Cumulative fluoro time	s	1
EF14	Acquisition frequency	$1/s$	0.476
EF4	Cumulative cine time	s	0.345
OF29	Long. position table	μm	0.227
OF30	Patient age range	–	0.217
OF9	Object thickness	cm	0.156
EF19	Long. distance travelled table	μm	0.139
OF22	Detector position	μm	0.098
OF28	Lat. position table	μm	0.083
OF24	Long. position frontal beam	μm	0.078
OF18	Rotation start	$^\circ$	0.076
OF19	Rotation end	$^\circ$	0.074
OF20	SID	m	0.073
EF5	Cumulative fluoro time	s	0.073
OF26	Frontal z-rotation	$^\circ$	0.065
OF23	Propeller position	μm	0.064
OF4	Run	kV	0.064
OF7	Cumulative acquisition time	s	0.063
EF18	Number of lat. moves table	–	0.059

3.4. Classification Model Overview

The final set of features, to be used as input for the random forest classification model in order to predict CAG procedure duration class, is presented in Section 3.3. Considering that the feature set and therefore the width of the input dataset has changed with respect to the initial dataset consisting of 52 features, the hyperparameters must be reevaluated. This is because the hyperparameters depend on the characteristics of the input dataset and affect model performance. The renewed optimal characteristics of the random forest classification model are found using the same steps as described in Section 3.2.3 and are presented in Table 3.8. With respect to the hyperparameters of the initial dataset, only the number of trees has changed from 300 to 350, in order to obtain the highest prediction accuracy.

Table 3.8: An overview of the hyperparameters of the final classification model.

Parameter name	Value
Number of trees	350
In-bag fraction	1
Number of features to sample	5
Maximum number of splits	14,017
Minimum leaf size	1
Minimum parent size	2

Now that the feature set is optimised and all parameters of the classification model have been decided upon, the model is ready for use. The results of the applied classification model are presented in Chapter 4.

3.5. Model Application Setup

Once the general model accuracy has been determined, the feasibility of the practical implementation of the model must be evaluated. Unfortunately, use of a schedule that corresponds to the dates from the available interventional X-ray data is not an option, since cath lab schedules are disposed immediately after completion of the procedures. Nevertheless, the available data can be used to analyse and interpret the CAG schedule retrospectively. Besides looking at application of the model with respect to the daily schedule, more insight is required with regards to individual procedures as well. The second part of the model application section analyses the timing of the predictions throughout the procedure, in order to assess its potential applicability. For example, the model would be considered more successful if it could predict the correct duration class early on in the procedure, rather than at the end of it.

For the retrospective interpretation of the schedule, a full day at the cath lab is analysed. To do so, a random day is selected during which five or more CAGs have been performed. This is to increase the chances of obtaining a dataset with CAGs of variable duration. The retrospective timetable is analysed and several procedures are studied individually. For a good understanding of the extent to which CAGs can be scheduled, the average duration of each CAG phase is provided in Table 3.9. According to these values, defined by K.M. van der Graaf, an entire CAG including turnaround takes approximately 48.6 minutes [28]. The definition of each phase can be found in Section 3.1.5. The Reinier de Graaf hospital assumes a CAG duration of 45 minutes for the creation of a cath lab schedule. Based on the provided average duration, break times and the fact that the cath lab requires preparation and cleaning, a maximum of nine CAG procedures can be scheduled per day. Within the available dataset, most days consist of three to four CAG procedures. The minimum and maximum number of CAGs per day corresponds to zero and ten respectively.

Table 3.9: The mean duration of the main phases of a CAG [28].

Phase of the CAG	Mean duration (minutes)
Patient preparation phase	11.8
Operative phase	22.4
Post-care phase	4.6
Turnaround	9.8
Total	48.6

In order to simulate an appropriate scenario for the analysis of individual procedures, the model is trained using all available data except for that one specific procedure. Subsequently, the model is tested on this procedure. This ensures that none of the acquisitions from a particular procedure have been used to train the model and examines the effect of this with respect to previous validations. The analysis should give an insight into the throughput of CAGs per day, as well as the duration of each and the time between each CAG. Furthermore, individual analyses provide a better understanding of the shifts between different duration classes throughout a procedure.

3.6. Feature Analysis Setup

The main goal of this research is to find the usefulness of the data from the interventional X-ray system for predicting total procedure duration. The final step of this analysis is to assess the added value of specific features that are logged during the procedure. Gaining insight into the relevance of particular features is beneficial for knowing which features could be used to detect delay and analyse workflow. Specifying such features allows for optimisation of the model and knowing which aspects of the data are relevant for future research.

Based on the 19 features that make up the columns of the final dataset, a feature importance analysis can be done using the *OOBPermutedPredictorDeltaError* function in Matlab. Using this function, Matlab reorders one column (feature) of the out-of-bag data for each tree and leaves the rest of the out-of-bag data the same. If the accuracy of the model performance on the out-of-bag data drops significantly, the feature is relevant for predicting procedure duration class. However, if there is no significant change in model accuracy, the feature is not particularly relevant for the model. This procedure is then repeated for all features belonging to each out-of-bag data subset, where each data subset corresponds to one decision tree from the random forest. In order to come to a final conclusion, the significance of each feature is deduced by averaging the importance of identical features from all decision trees.

Once the importance of all features has been rated, the individual features can be analysed in more detail. For the five most important features, three different types of analyses are performed. Initially, the range of values within each feature is analysed, providing an indication of variance. Thereafter, an analysis is done to investigate whether the value of the feature changes with respect to different procedure duration classes. This is especially useful for determining the degree to which the feature is useful to analyse procedure duration and workflow. Finally, the behaviour of the each feature throughout the procedure is analysed, where each duration class is presented separately. The analysis is valuable to detect whether the feature values change as the procedure progresses and whether this changes with respect to procedures of varying total duration.

4. Results & Discussion

This chapter presents the results obtained from running the optimised dataset from the interventional X-ray system in a fine-tuned random forest classification model. Initially, the ability of the model to classify the data points and predict the approximate procedure duration is presented in Section 4.1. The results are discussed in Section 4.2. Subsequently, the feature importance of the finalised feature set is presented in Section 4.3 and the five most important features for predicting procedure duration are presented in Section 4.4. Finally, the interpretations of the results are discussed in Section 4.5.

4.1. Classification Accuracy

The ability of the model to estimate procedure duration has been tested on two different types of datasets, to ensure in-depth validation of the data usability. Section 4.1.1 presents the results of testing the model on out-of-bag data, while Section 4.1.2 presents the cross-validated results.

4.1.1. Out-of-bag Accuracy

The out-of-bag dataset of each decision tree that makes up the random forest, represents the data that was not part of the subset used to generate the decision tree. Based on the hyperparameter tuning methods explained in Section 3.2.3, a dataset of the same size as the original data is used. Nevertheless, since each data point is selected using replacement, there are numerous repetitions in the dataset. All data points that have not been used to train the model make up the out-of-bag data. This is information that the decision tree has never seen before and can therefore be used to evaluate how it would perform on new data. When running the model on each decision tree using its respective out-of-bag dataset and averaging the error, the general performance is found. For this specific dataset and model, the out-of-bag misclassification probability is 7.25%. This implies that 92.75% of all acquisitions have been correctly assigned to the duration class of their corresponding procedure. The classification accuracy per class can be presented by means of a confusion matrix, as shown in Figure 4.1.

True class (minutes)	0-10	7029	135	1		98.1%	1.9%
	10-20	432	3823	4		89.8%	10.2%
	20-30	117	81	1386	1	87.4%	12.6%
	>30	42	63	1	903	89.5%	10.5%
		92.2%	93.2%	99.6%	99.9%		
		7.8%	6.8%	0.4%	0.1%		
		0-10	10-20	20-30	>30		
		Predicted class (minutes)					

Figure 4.1: Confusion matrix of the predicted versus the true duration based on out-of-bag data.

The confusion matrix in Figure 4.1 shows the duration predictions of all acquisitions in the out-of-bag data versus their actual duration class. The blue colour represents the degree to which acquisitions are accurately estimated. Red tones represent the degree to which the acquisitions are incorrectly classified. In both cases, a darker colour refers to a higher accuracy and error respectively. The horizontal error bar shows the ratios of correct and incorrect estimations of procedure duration per predicted class. Vertically, the percentages of correct and incorrect classifications within the actual duration classes are shown. According to the figure, classification errors are mostly due to underestimation of procedure duration, rather than overestimation. For only 141 out of 14,022 acquisitions, the procedure is estimated to take longer than it actually does. For 736 acquisitions, the total procedure duration is estimated to be lower than its true value. The duration class with the most incorrect classifications is the 10 to 20 minute class, for which 434 acquisitions are misclassified. However, with respect to the total number of acquisitions per class, the duration class with the highest error rate is the 20 to 30 minutes time frame, with a misclassification rate of 12.6%. In comparison to the three highest procedure duration classes, the misclassification rate within the class that represents a procedure duration below 10 minutes is rather low, namely 1.9%. Looking at the horizontal error bar, it seems that more than 90% of the predictions made for all duration classes are correct. Predictions that state that the procedure will take 20 minutes or longer are correct over 99% of the time. Given the fact that the interventional X-ray data is not made for the purpose of estimating procedure duration and mostly logs technical parameters regarding machine functionality, the results present an unexpectedly high accuracy.

4.1.2. Cross-validated Accuracy

A different method of validating the classification results is by running the optimised random forest model using test data. This is data that was excluded during the initial model creation and has never been seen by the model before. Based on this validation method, the resulting misclassification probability equals 6.33%. The corresponding confusion matrix can be seen in Figure 4.2.

True class (minutes)	0-10	1249	24	1		98.0%	2.0%
	10-20	83	694			89.3%	10.7%
	20-30	15	8	246		91.4%	8.6%
	>30	4	8		141	92.2%	7.8%
		92.5%	94.6%	99.6%	100.0%		
		7.5%	5.4%	0.4%			
		0-10	10-20	20-30	>30		
		Predicted class (minutes)					

Figure 4.2: Confusion matrix of the predicted versus the true duration based on test data.

Whereas the data in Section 4.1.1 consists of 14,022 acquisitions, the test data has only 2,473 data points. The cross-validated error rates are lower than the out-of-bag error rates, especially for procedures with a duration of 20 minutes or longer. Only the error rate for procedures that have taken between 10 and 20 minutes is slightly higher based on cross-validation, namely 10.7% with respect to 10.2% for the out-of-bag data. Again, more procedures are underestimated rather than overestimated in terms of duration. In fact, only procedures that took under 10 minutes have been estimated to take longer.

For procedures with an actual duration of 10 to 30 minutes, duration was never estimated to be higher than its true value. All predictions that state a duration of 30 minutes or longer are correct. Out of the 247 predictions that estimate a duration of 20 to 30 minutes, only one appears to be incorrect. This implies that predictions by the model stating a duration of 20 minutes or over are particularly accurate. The downside is that as of yet, it is uncertain at which point this prediction occurs. If the model predicts the procedure to take longer than 20 minutes when 19 minutes of the procedure have passed, the additional value of the prediction is relatively small. In order to analyse this further, the classification timing is analysed in Section 4.1.3. Furthermore, individual interventions are analysed in Section 4.1.4. In order to minimise bias within further results and use the biggest error to evaluate the classification model, all further analyses are performed based on out-of-bag data.

4.1.3. Classification Timing

For understanding the functionality of the model, it is essential to analyse its classification abilities beyond the average accuracy. One way of gaining insight into its predictive abilities throughout the procedure, is by plotting the predicted class versus the cumulative procedure time. By doing this for every class, information can be collected regarding changes in predicted duration class throughout the procedure and how this varies among the duration classes. Figure 4.3, 4.4, 4.5 and 4.6 present the predicted duration for each acquisition, separately graphed for every duration category. The graphs give a general insight into the occurrence and timing of predictions for each duration class. For a more detailed insight into the point at which predictions change to a different duration class, individual procedures have to be analysed. This is realised and further elaborated on in Section 4.1.4.

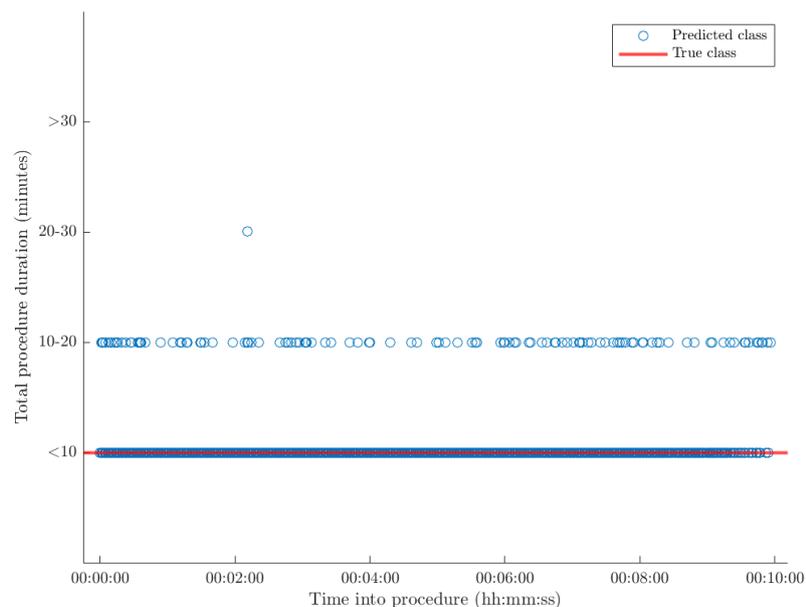


Figure 4.3: Procedure duration predictions throughout all procedures that took 0-10 minutes.

For all acquisitions that were obtained from procedures with a total duration of 10 minutes or less, 98.1% was classified correctly. Figure 4.3 shows that the rate at which acquisitions are incorrectly classified as Class 2 does not decrease throughout the procedure. This could have to do with the fact that towards the end of the procedure, indications that the procedure will exceed the 10-minute margin go up.

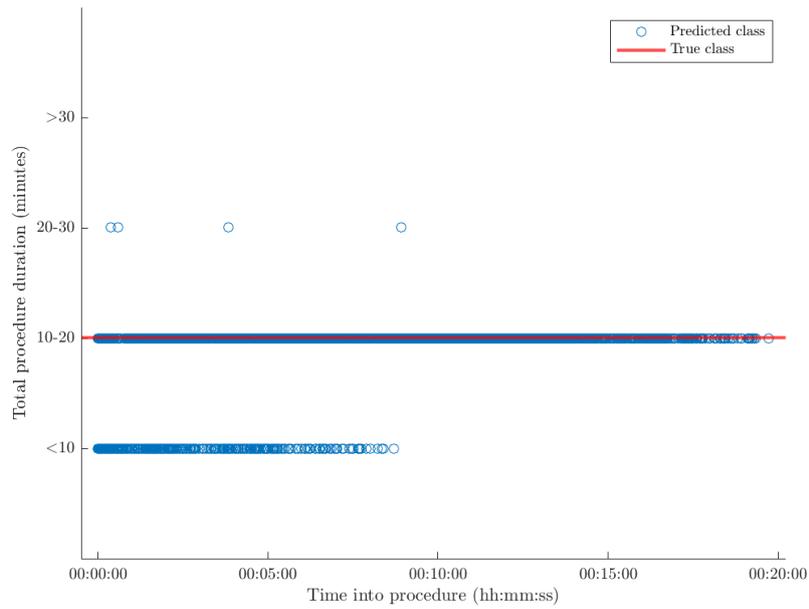


Figure 4.4: Procedure duration predictions throughout all procedures that took 10-20 minutes.

Figure 4.4 shows predictions for the second duration class. As the cumulative procedure time approaches the 10-minute margin, predictions that incorrectly estimate a duration of less than 10 minutes gradually decrease. Nevertheless, underestimation appears to occur until the margin itself. Overall, the predictive accuracy within the procedures that last between 10 to 20 minutes is 89.3%, as shown by Figure 4.1. This is confirmed by the fact that the vast majority of the data points are plotted on the correct line representing true duration class. Unlike Figure 4.3, Figure 4.4 does not show misclassification until the end of the procedure. As a matter of fact, no incorrect predictions are made after the 10-minute margin, not for higher duration classes nor for lower ones.

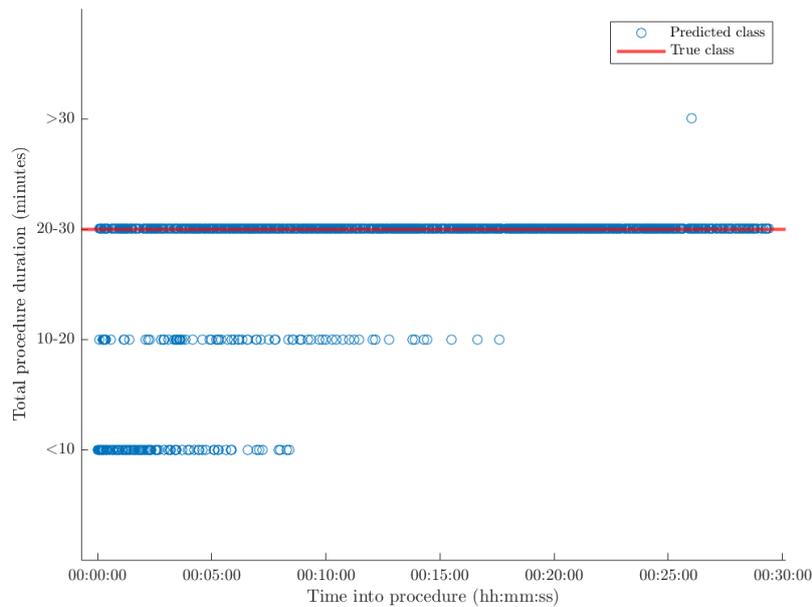


Figure 4.5: Procedure duration predictions throughout all procedures that took 20-30 minutes.

In Figure 4.5, all duration predictions for procedures with a total duration of 20 to 30 minutes are shown. In some cases, the correct duration class is estimated from the first minute onward. In other cases, the model first predicts the total duration to be 0 to 10, or 10 to 20 minutes. The latter predictions occur until the end of their respective margins. Ideally, this would not be the case, as being able to make accurate, unforeseen predictions would increase the relevance of the model. Nevertheless, this graph does not show a distribution of predictions for individual procedures, thus the current information is not enough to draw such conclusions yet. According to Figure 4.1, this duration class shows the highest error rate, namely 12.6%. This implies a prediction accuracy of 87.4%.

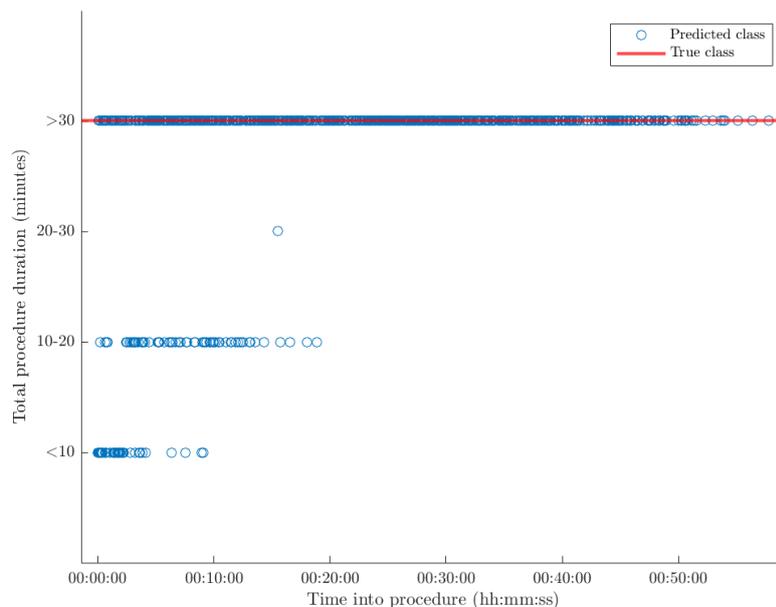


Figure 4.6: Procedure duration predictions throughout all procedures that took over 30 minutes.

Predictions for a duration of 30 minutes or more, as shown in Figure 4.6, appear to be more useful than the distribution shown for the third duration class. From the 20-minute margin onwards, no incorrect predictions are made. This implies that for all acquisitions done past the 20-minute margin, the prediction that the procedure will take over 30 minutes is correct. For clinical application of the model, this is useful for surgical planning. The accuracy of predictions for procedures from the last duration class is 89.5%, as presented by Figure 4.1.

4.1.4. Model Application

In order to analyse the model in further depth and evaluate the feasibility of practical application, operating days and procedures are analysed individually. Initially, a random day within the dataset is selected in order to assess the duration of each procedure, the throughput of patients and the change in predicted class with respect to progress. Subsequently, several procedures are analysed that stand out positively or negatively in terms of predicted duration. On top of the data from the interventional X-ray system, DoseWise data is used to register the time that the patient documents were uploaded. This gives a better indication of the total time of the procedure, including patient preparation time. The parameter is relatively important for analysing the amount of time each patient occupied the cath lab, since this is not registered by the interventional X-ray system. The randomly selected day that is analysed in further depth is the 25th of February 2019, during which seven CAGs were performed.

February 25th 2019

On this day, the CAGs that took place in the cath lab of the Reinier de Graaf hospital are shown in Table 4.1. The starting time, ending time and operative duration are logged by the interventional X-ray system. The time that the patient file was uploaded onto DoseWise is shown in the third column. This value can be used as an indication of when the patient entered the cath lab [28]. The operative duration is the time between the first and the last acquisition, whereas total duration refers to the duration between the patient document upload time and the last acquisition. The turnaround time of a procedure is the time between the end of the previous procedure and the DoseWise file upload time of the current patient.

Table 4.1: All CAGs that took place in the cath lab of on February 25th 2019.

Exam number	Turnaround time	DoseWise upload	Starting time	Ending time	Operative duration	Total duration
4181	-	09:06:00	09:28:28	09:38:57	00:10:29	00:32:57
4182	00:19:03	09:58:00	10:13:06	10:29:43	00:16:37	00:31:43
4183	00:25:17	10:55:00	11:00:13	11:57:59	00:57:46	01:02:59
4184	00:29:01	12:27:00	12:37:05	13:12:43	00:35:38	00:45:43
4186	01:02:17	14:15:00	14:29:28	14:42:03	00:12:35	00:27:03
4187	00:28:57	15:11:00	15:14:38	15:31:21	00:16:43	00:20:21
4188	00:27:39	15:59:00	16:00:11	16:13:10	00:12:59	00:14:10

Table 4.1 shows that the CAGs performed on this day show a high variability in terms of procedure duration. The first procedures seem to run smoothly, with a total duration that is below the average value of 48.6 minutes and the predetermined scheduled duration of 45 minutes, excluding post-care time. However, the operative duration of the procedures with exam number 4183 and 4184 are above average. The three follow-up procedures in the afternoon are of a relatively short duration. The procedures analysed in this section include exam number 4181, 4183 and 4184.

The predictions for the procedure with exam number 4181, which has an operative duration of 10 minutes and 29 seconds, are shown in Figure 4.7. Up to approximately 2 minutes, the model predicts the correct duration class, until the prediction changes to Class 1. From the fifth minute onwards, the output values go back to representing the correct duration class. From earlier analysis, it is known that for procedures that take 0 to 10 minutes, the class can be overestimated up to the 10-minute margin. In this case, the procedure took just over 10 minutes and the model is correct in its predictions. However, the model does not provide much additional value in terms of an analysis of the workflow and an estimation of procedure duration. Several acquisitions show an underpredicted duration. Moreover, the degree to which the predicted duration of 10 to 20 minutes is useful, is questionable for this specific procedure. With a duration that exceeds the first class by 29 seconds, the estimation that states a 10 to 20 minute duration might be misleading. For procedures of such short duration, the 10-minute intervals that define each class might be too big and not representative enough. An solution would be to implement smaller classes for procedures with a duration below 10 or 20 minutes. A logarithmic scale could be useful in this case, where class size increases as duration prevalence decreases.

The red and green lines represent the last and first predictions of a newly predicted duration class respectively. This is done to analyse the effect of using either prediction type as a reference point and to determine the one that is more reliable. For this specific procedure, neither indications are of much additional value.

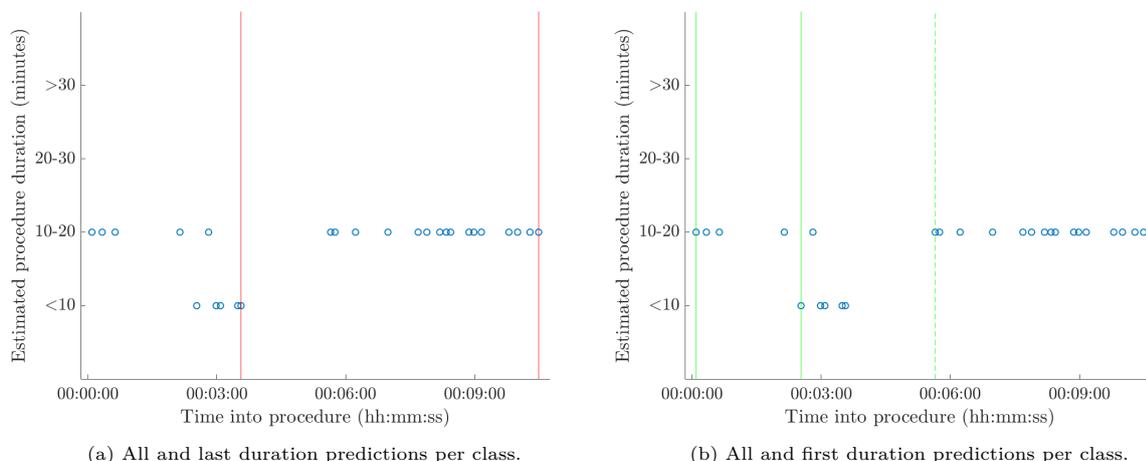


Figure 4.7: All procedure duration predictions throughout procedure 4181.

The next analysed procedure has a total operative duration of 57 minutes and 46 seconds. The classification model's predictions for this procedure are shown in Figure 4.8. Whereas the previous procedure seemed to show overlap in duration class predictions, this graph shows a clear upward pattern without any overlap. Using the last prediction for each duration class results in predictions as shown in Figure 4.8a, which is not much different than using the first predictions as shown in Figure 4.8b. Interesting to note is that the procedure is predicted to take longer than 10 minutes from the first acquisition onward. This means that particular starting values of the registered features within this dataset are indicative of a total duration that is longer than average. At all stages, the model is useful for detecting that the procedure will take longer than usual. Within the first 10 minutes, the model predicts the procedure to take longer than 20 minutes. The shift to the fourth duration class is slightly less valuable, as this is done just before the 30-minute margin. If this model were to be used for predicting procedural duration more accurately than the physician in the room, it is not exceptionally useful. However, if the model would be used to update a third party on the status and the workflow of the procedure, the model can definitely be considered useful.

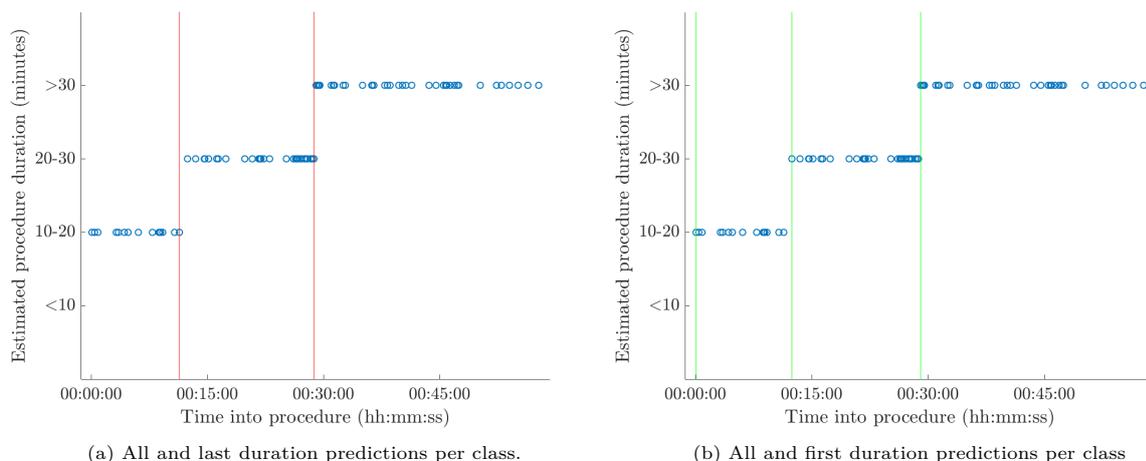


Figure 4.8: All procedure duration predictions throughout procedure 4183.

Figure 4.9 shows the predictions for procedure 4184, which took 35 minutes and 38 seconds in terms of the operative duration. The graph shows that the duration is never predicted to be lower than 10 minutes.

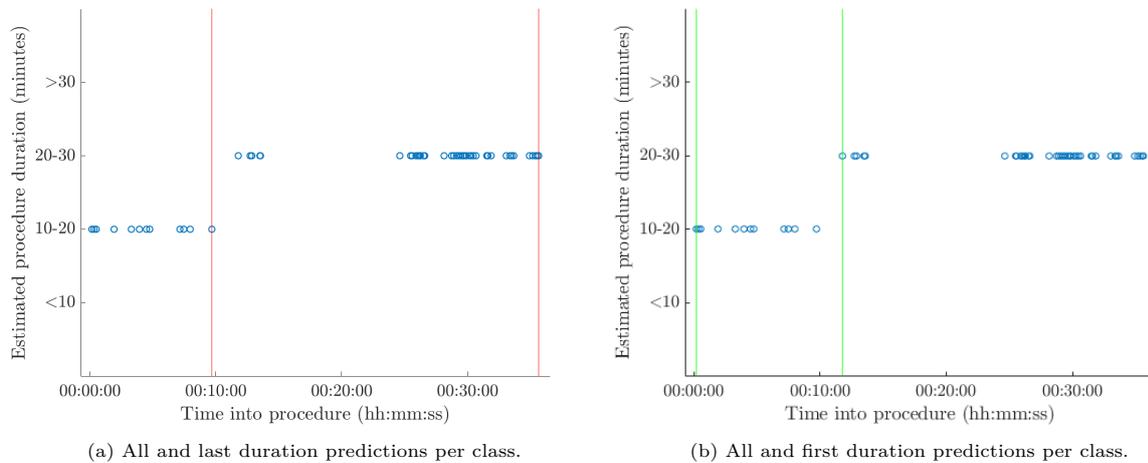


Figure 4.9: All procedure duration predictions throughout procedure 4184.

Strangely enough, the model never predicts the correct duration class for this procedure. Even beyond a cumulative procedure time of 30 minutes, it predicts a duration of 20 to 30 minutes. Section 4.1.3 shows that for the original model, this was never the case. Perhaps this occurs because the entire procedure was left out of the training model. This would imply that in the general classification analysis in Section 4.1, similar acquisitions from the same patient could have biased the model and therefore prevented such errors. In order to improve the current model, a conditional restriction should be added. When preventing the predicted duration class from being lower than cumulative procedure duration, the graph would look as follows.

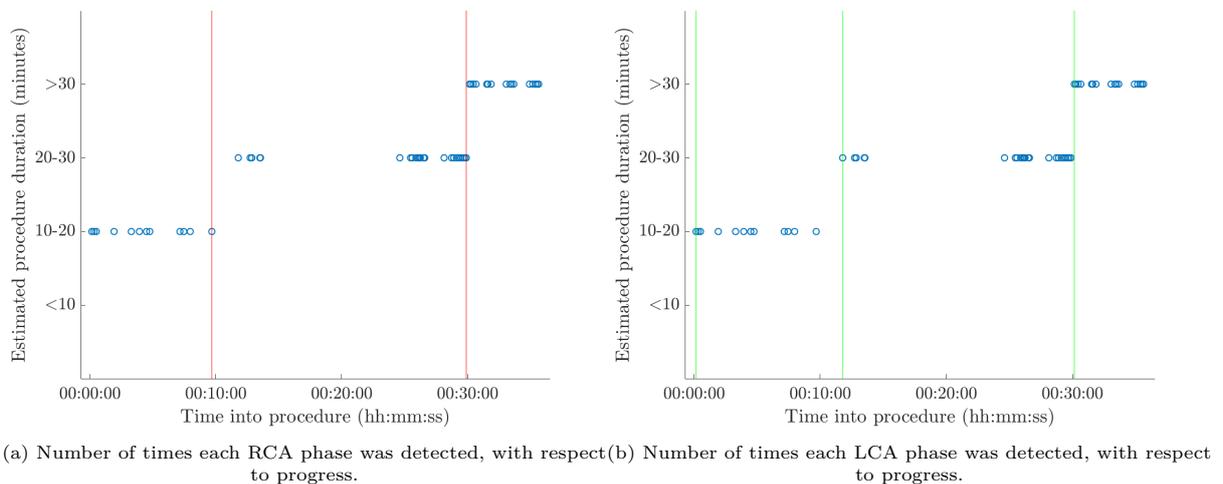


Figure 4.10: All renewed procedure duration predictions throughout procedure 4184.

Given the fact that the model might predict a duration that is lower than the cumulative procedure time, the restriction shown in Figure 4.10 should be included in the final model, for backup purposes. The effect of repeated measures, which is a possible reason for why this has not been seen in previous testing, is further discussed in Section 4.2.

May 22nd 2018

For procedures under 10 minutes, results have shown that the model tends to overestimate the total duration by predicting a Class 2 duration. Procedure 3221, which took a total operating time of 6 minutes and 3 seconds, was predicted to take under 10 minutes during the first few acquisitions, as can be seen in Figure 4.11. This indicates that the model does differentiate between procedures that are likely to take 0 to 10 minutes and procedures that might take more time from the very start. However, before the 2-minute mark, the prediction has already shifted to the second duration class. This verifies that the assumption that a newly predicted duration class is always reliable, does not apply to predictions under 10 minutes. Only after 10 minutes have passed can such assumptions be used.

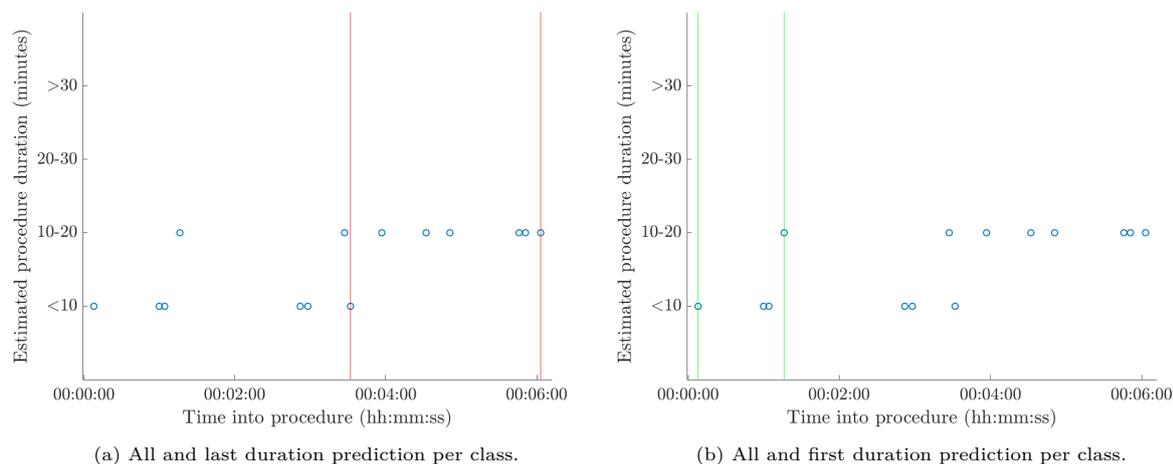


Figure 4.11: All procedure duration predictions throughout procedure 3221.

April 20th 2018

An example of a successful application of the model is shown in Figure 4.12. This figure shows that when the first prediction of each class is taken as the principle indication of duration class, as depicted by the green vertical lines, accurate predictions are given early on in the process. At approximately 10 minutes, it states that the procedure will take over 20 minutes and at 15 minutes into the procedure, it knows it will take longer than 30 minutes. Previous results have shown that after 10 minutes, the first prediction of a new class can be seen as an accurate estimation of duration. Generally, the green lines can be considered appropriate reference points, in particular for procedures over 10 minutes. After this time, duration is rarely overestimated.

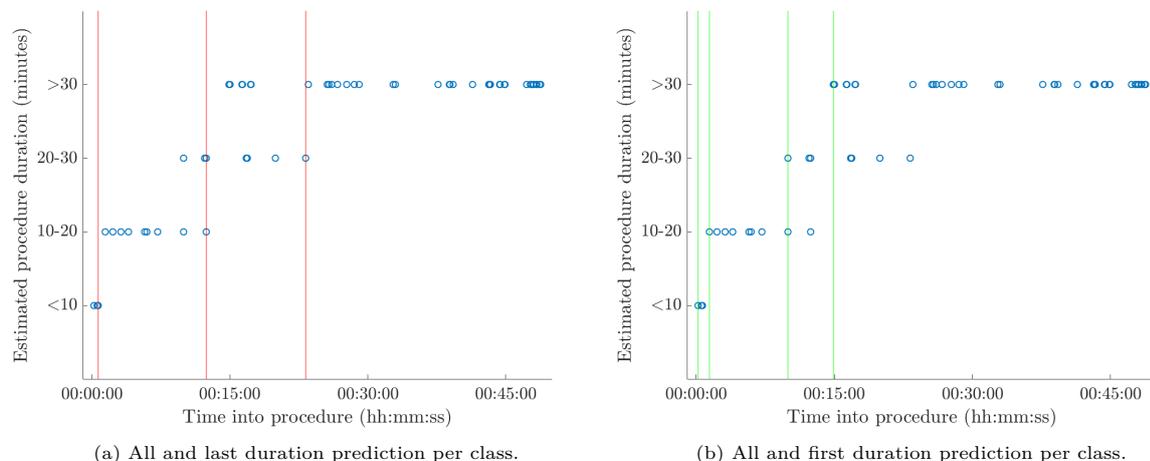


Figure 4.12: All procedure duration predictions throughout procedure 3124.

General applicability

In order to assess the overall ability of the model to classify each phase at a useful point throughout the procedure, general overviews are made. Each general overview below shows a distribution of the time at which the particular class was predicted for the first time, corresponding to the green lines in prior figures.

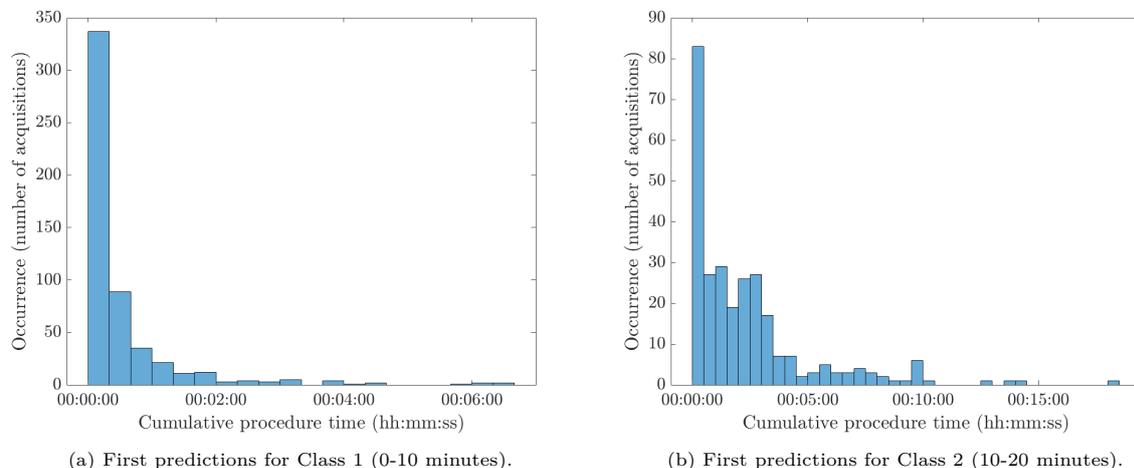


Figure 4.13: The procedure times at which the first predictions for Class 1 and Class 2 are made.

The first graph is not of the same relevance as the other three, as it has been proven that the predictive accuracy is significantly lower for procedures of a shorter duration. Nevertheless, once the cumulative time has exceeded the 10-minute margin, predictions are reasonably accurate. Figure 4.13a shows that Class 1 is mostly predicted within the first 2 minutes of the procedure, especially during the first 30 seconds. According to Figure 4.13b, Class 2 is mostly predicted during the first 4 minutes, yet again mostly during the first 30 seconds. Based on the classification matrices, inaccuracies in the predictions among these two classes are relatively rare. Nevertheless, the errors are noteworthy and as of yet, the predictions made during the first few minutes of a procedure are not accurate enough to be used for important application.

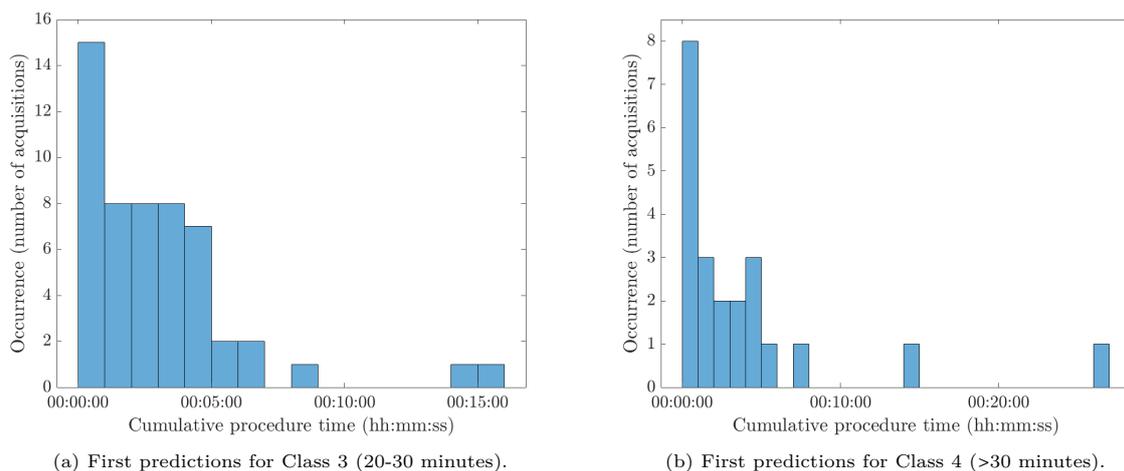


Figure 4.14: The procedure times at which the first predictions for Class 3 and Class 4 are made.

Figure 4.14 shows that predictions for the third and fourth class often occur relatively early on in the procedure. In Figure 4.14a, the majority of the predictions are made in the first 5 minutes, with a peak in the first minute. Almost all of the significant delays, which refers to a procedure duration between 20 and 30 minutes, are detected at least 10 minutes before the cumulative procedure time hits the 20-minute mark. The fact that for 15 different procedures, a minimal duration of 20 minutes was predicted in the first minute proves the promising predictive ability of the interventional X-ray data. The distribution of the fourth duration class, as shown in Figure 4.14b, also shows a peak in the first minute. Generally, the predictions for the fourth class occur within the first 10 minutes of a procedure. This means that most of the severe delays are detected at least 20 minutes in advance. Again, it shows the ability of the model to predict large delays at a useful point in the procedure.

4.2. Interpretation of the Classification Results

The application of the optimised random forest model to the available data has resulted in a classification accuracy of 92.75%. Given the fact that the technical data is not originally made for the purpose of progress analysis, this is remarkably high. The random forest model without the additional features from Section 3.1.4 and without RFE optimisation gives a classification accuracy of 88.24%. This shows that the original data shows suitable characteristics, but that additional features and fine-tuning of the feature set have made a significant contribution. The out-of-bag classification accuracy for Class 1, Class 2, Class 3 and Class 4 procedures was found to be 98.1%, 89.8%, 87.4% and 89.5% respectively.

For further analysis of the functionality of the classification model, acquisitions from every procedure class were evaluated with respect to the time at which the correct duration was predicted. This also provides insight into how predictions change with respect to progress. For particular Class 1 procedures, the classification model estimated a duration of 10 to 20 minutes, which occurred up to the 10-minute margin. However, only 1.9% of the procedures with a duration below 10 minutes were classified incorrectly. For procedures that took between 10 and 30 minutes, incorrect predictions were made up until the starting margin of the correct duration class. For example, several acquisitions from Class 2 procedures, which take between 10 and 20 minutes, were incorrectly classified up to the 10-minute margin. The closer the predictions are to this margin, the less relevant they are for predicting duration. At 9 minutes into the procedure, it is less useful to learn that the procedure will take more than 10 minutes than if this were told in the first few minutes. Therefore, earlier prediction is desirable. For procedures that take 30 minutes or more, incorrect estimations are made up to the 20-minute margin. This means that predictions for this duration class can generally be considered more useful, as they are correctly predicted earlier on with respect to the overall duration of the procedure. Furthermore, Class 1 acquisitions showed to be the only data points for which procedure duration was overestimated. This is beneficial, as it means that when Class 3 or Class 4 are predicted, this is most certainly the minimum duration class of the corresponding procedure. Also, it means that predictions beyond the 10-minute margin can be considered reliable indicators of minimum procedure duration.

A limitation of the aforementioned graphs is that they provide no insight into the density of acquisitions with respect to cumulative procedure time. For multiple acquisitions within a specific time frame, the data points simply overlap. Considering the minimum classification accuracy of 87.4%, the graphs give a slightly pessimistic perspective of the model's effectiveness, as the incorrectly classified acquisitions stand out most because of less overlap. Further limitations of the graphs include the fact that the distribution of acquisitions from individual procedures are not taken into account. Moreover, the general distribution of acquisitions throughout all procedures, as shown in Appendix A, is not considered either. For example, fewer acquisitions are generally applied in the first half of the procedure with respect to the second half. In the graphs from Section 4.1.3, the general absence of observations at certain points throughout the procedure might be misleading and suggest classification errors. Furthermore, the procedures are categorised based on 10-minute time slots, whereas the total duration of procedures varies within this time frame. Naturally, the number of acquisitions decreases as the cumulative procedure time increases, as fewer procedures last for this amount of time.

All of the aforementioned limitations are factors that were implemented for a more representative presentation of the results, as shown in Section 4.1.4. Within this section, model applicability was tested by analysing various individual procedures from a randomly selected day. Within this day, the operative duration ranges from 10 minutes and 29 seconds to 57 minutes and 46 seconds. For the shortest procedure, the predicted duration classes include Class 1 and Class 2. Since the procedure duration barely exceeds the 10-minute margin, the additional value of the prediction stating a duration of 10 to 20 minutes is questionable. This proves that smaller duration classes would benefit the model. For delayed procedures, the model seems to detect a duration of at least 20 minutes from the first acquisition onwards. This shows that the values of certain features can indicate a difficult case or a delay early on in the process. However, the exact delay is generally only estimated later on in the process. For one of the analysed Class 4 procedures, the procedure was predicted to take 20 to 30 minutes within the first 10 minutes.

At 15 minutes into the procedure, its duration was estimated to be over 30 minutes. This shows that a duration analysis using a random forest algorithm is capable of providing early information regarding delays, revealing promising scheduling benefits. For accurate prediction beyond 30 minutes, additional classes would have to be added to the model.

In the analysis of individual procedures shown in Section 4.1.4, a comparison is made between using the first time at which a duration class is newly predicted and the last time this occurs, depicted by green and red vertical lines respectively. Based on the aforementioned conclusions, first-time class predictions can be considered reliable beyond the 10-minute margin, or when it concerns a Class 3 or Class 4 prediction. The individual analyses confirm this statement. For short procedures, the use of either reference point is suboptimal, due to fluctuations between predictions that state a Class 1 and a Class 2 duration. Nevertheless, for long procedures, the model successfully detects delays, resulting in early detection when the first-time predictions are used. Further analysis of the first-time detection of new classes show that Class 3 and Class 4 procedures are normally detected within the first five minutes of a procedure. Therefore, the model proves to be extremely successful in detecting significant delay at an early point throughout the procedure.

Regarding the additional value of the model to procedure scheduling, it could reduce waiting time of the next patient and improve general planning. For the specific day analysed in Section 4.1.4, accurate predictions could have decreased the amount of time in between procedures that were done in the afternoon. Nevertheless, the retrospective schedule provides no information on break times or why the cath lab was empty for a relatively long period of time between different CAG procedures. Therefore, no definite conclusions can be drawn regarding how the application of the model would have affected this day, since the suggested improvements are merely based on assumptions. Further information is required to validate the applicability. This is a limitation of the applicability analysis, which could be further analysed through real-life observation or by means of individual patients files. Nevertheless, the classification model has shown valuable predictive abilities regarding the detection of delay. Beyond the 10-minute margin, it is able to provide an accurate representation of workflow and procedure duration is generally not overestimated. The intraoperative predictions could serve as a useful indication of workflow for third parties, such as schedulers. Furthermore, they could serve as an indication for when to call the next patient. By knowing that the procedure is delayed, one can wait to call the new patient or to schedule extra procedures. Nevertheless, to ensure optimal patient throughput, the 10-minute classes could be reduced to 5-minute slots or less, in order to increase prediction precision. This is because procedures might still be completed earlier within the 10-minute time frame of the predicted duration class. Such changes would need to be implemented before the model is suitable for application in clinical settings. Only 11.52% of the procedures within the used dataset are delayed beyond 20 minutes. In order to make the model financially viable, further refinement of all classes is needed. This would ensure that the model is not only able to detect delay, but can also specify the delay more accurately and predict the duration of shorter procedures. The main difficulty in predicting the duration for shorter procedures is that at the start of a procedure, there is still a lot of uncertainty regarding procedural events. As more acquisitions are made, more insight into the procedure is gained. Further research could show whether more conclusions can be drawn from the initial acquisitions. Alternatively, it would be interesting to implement more data that does not vary throughout the procedure, such as patient and staff characteristics.

Out-of-bag vs. test data

The evaluation of classification accuracy was based on two different types of validation methods and provides an interesting perspective on how the applied method can affect the results. Out-of-bag validation is a suitable method for using the dataset to its fullest potential and is appropriate for small datasets. This is because all training data is drawn with replacement, implying that an infinitely large training set can be generated. Therefore, it is particularly appropriate for the implementation of data from the third and fourth duration class, which contain a relatively small number of samples. However, the quality of the training set reduces with excessive replacement. For out-of-bag analysis, the accuracy is based on results from individual trees, using variable validation data. This ensures that dominant values and outliers are included, but not in every tree, creating a generalised model with lower chances of overfitting.

For cross-validation, the classification model uses one set of validation data. The cross-validation test data might contain fewer outliers, which could explain why the cross-validated error is lower. It was important to implement cross-validation on top of out-of-bag validation, since the model was optimised using hyperparameter tuning based on the out-of-bag classification error. However, since the out-of-bag error appeared to be higher than the cross-validated error, it can be assumed that this did not bias the classification model. In order to reduce optimistic bias, all predictions made beyond the evaluation of general classification accuracy are made and evaluated based on out-of-bag data. Both methods are applied to an identical model that only used 85% of the original dataset to train the classification model. Therefore, they both provide a slightly more pessimistic view of the classification abilities of the interventional X-ray data than if the entire dataset were to be used as training data.

Repeated measures

The analysis of individual procedures in Section 4.1.4 provide an interesting insight into the applicability of the model. For each individual analysis, a new model was trained using all data except for the acquisitions from the analysed procedure. At the start of the project, the fact that random forest does not take into account data from previous acquisitions of the same patient was seen as a limitation. However, the exclusion of entire procedures from the training data for the analysis of individual procedures has shown that the overlap between repeated measures might be bigger than originally anticipated. This is based on the fact that the individual procedures show phenomena that were not present in the overall classification analysis, such as class predictions that suggest a lower total duration than the corresponding cumulative procedure time of the acquisition. The data points within the available dataset could be considered dependent because measurements are repeated on specific patients. For most features, this value changes for each acquisition, but the exact resemblance has not been verified. Therefore, the extent to which the repeated measures might have caused optimistic bias to the general classification accuracy in Section 4.1 is uncertain. By means of elimination of the repeated measures from the training data, chances of optimistic bias were reduced significantly. Despite the fact that classification accuracy seems slightly lower, the model still shows promising abilities in terms of detecting delay. This can be seen in Figure 4.8, Figure 4.9 and Figure 4.12. Furthermore, the first time Class 3 or Class 4 is predicted is still a valid reference point for minimum procedure duration. Future research is required to investigate the effect of repeated measures on the accuracy of the random forest classification model. For this research project, a random forest classification method was selected due to its excellent ability to analyse feature importance. Furthermore, it is suitable for dealing with datasets that include a high number of features. Nevertheless, other machine learning techniques that are more suitable for dealing with repeated measures, such as regression, could be used to study the effect of dependent data points.

Related work

Multiple studies have used machine learning to predict surgical duration, using various sources of data. Kargar et al. looked into three different machine learning methods to estimate procedure duration of 104 different procedures [49]. A random random forest regression algorithm was found to provide better results than linear regression and multivariate adaptive regression splines. The used data includes administrative and elective surgery data, such as patient details, surgery characteristics and the medical physicians involved. The resulting model improved the predictions made by current hospital methods by 28%. Nevertheless, the quality of the current predictions are unknown. Techniques implemented by other studies include ANOVA and linear mixed models [13, 54]. The studies showed an improvement of 15% and 36%, compared to estimations made by the physician. Most of the found studies focus on the effect of predetermined parameters on the duration, such as surgical team composition, experience and time of the day. Therefore, they focus on preoperative duration prediction, rather than using intraoperative data for real-time prediction. In order to enable the shift from a static to a dynamic schedule, intraoperative analysis is required. This is a shortcoming that has been addressed in this report, where online data is used to update duration predictions as the procedure progresses.

Besides estimating duration, some studies were found to predict the surgical workflow phase. Stauder et al. used random forest classification to intraoperatively predict the surgical phase of a laparoscopic cholecystectomy [53]. The data includes information regarding instrument usage and data such as the weight of the irrigation and suction bags. The model resulted in a classification accuracy of 68.78%.

A significant shortcoming of this study is that all surgical workflow phases had to be defined manually. This was done based on the recorded data points from only four procedures, by means of manual classification of each training sample. This is a considerable disadvantage of implementing phases that are not defined in terms of the input data. Appropriate methods are required to validate the phase, which is often not possible due to a lack of information of the labour intensity of the process. The research done in this report addresses this shortcoming by removing the need to implement phases, through evaluation of the workflow with respect to time. As this is an automatic process based on readily available information, the random forest model can be optimised indefinitely as the available input data increases. According to F. van Luyn, the amount of cases used in online models ranges from 1 to 80, due to the aforementioned reasons [56]. Nevertheless, the online model created in this report is based on 599 procedures.

No studies were found that use random forest classification to estimate duration. A likely reason is that regression seems more appropriate, generating a numerical value for the estimated remaining duration. Furthermore, most studies focus solely on improving the predicted duration, using static data that does not change throughout the procedure. For offline data, sensitivity to sudden changes is significantly smaller, which increases the likelihood of obtaining a regression model that provides enough accuracy for reliable conclusions. However, the aim of this report is to evaluate the usability of particular data that is updated intraoperatively. Considering the early stage of this study, the use of classification allows for more certainty of the drawn conclusions, regarding the quality of the data and the behaviour of specific features.

The next sections focus on the second part of this research, where the influence of various features on the predictions are analysed.

4.3. Feature Importance

Figure 4.15 presents the importance of each feature for classifying data points in terms of procedure duration. The feature set implemented in this analysis is the final feature set and consists of 19 features as deduced by the RFE in Section 3.3.

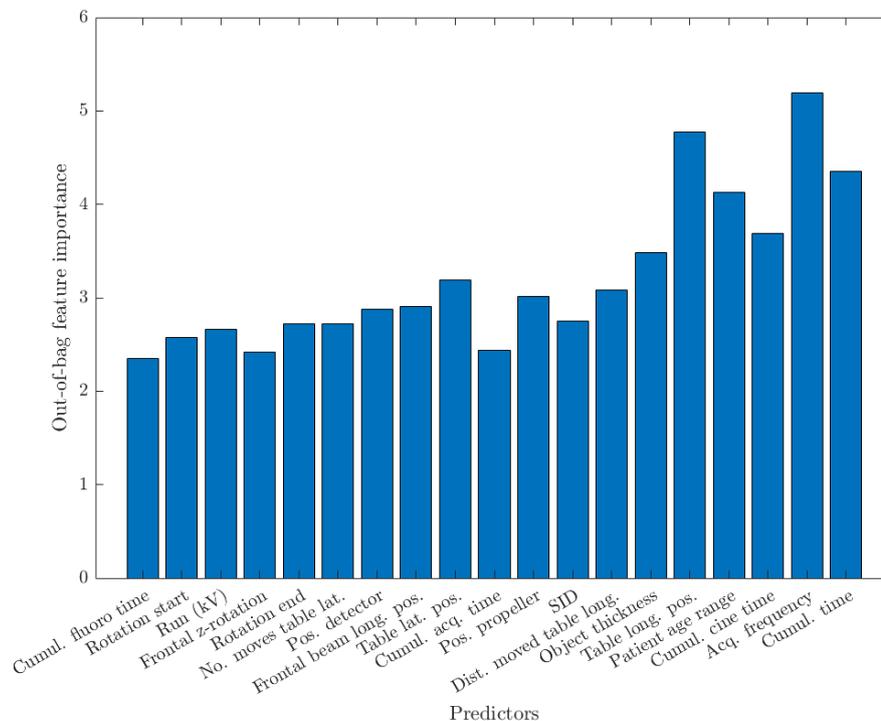


Figure 4.15: The out-of-bag importance of all features for the random forest classification model.

A definition of the out-of-bag feature importance can be found in Section 3.6. The features on the x-axis are presented in the order in which the features were eliminated in the RFE. Therefore, it could be argued that the importance should increase gradually as one moves to the right along the row of features. However, relative importance changes according to the input dataset and the graph shows that the feature importance for this specific feature set is indeed different. The most important feature for predicting procedure duration is acquisition frequency, which is the average number of acquisitions per unit time. The second most important feature is the longitudinal position of the table, thus the extent to which it longitudinally moves towards and away from the centre of the X-ray system. Other important features include cumulative procedure time, patient age and cumulative cine time. Here, cine refers to the selected acquisition method. Cine acquisitions are specifically used to for capturing and recording arteries, rather than guiding the catheter to the heart. The five most important features are analysed in more depth in Section 4.4.

Interestingly enough, no features that refer to the position of the X-ray system are among the most important features to predict total duration. However, the position of the table seems to be especially relevant, particularly the longitudinal position. The fact that multiple features refer to the position of the operating table means that this component from the cath lab must certainly be taken into account in further research. Besides only focusing on the angulation and rotation of the C-arm, as was done in Section 3.1.5, the operating table could also be used define the phases of a CAG. In fact, multiple features are suitable for further research into phase definitions and more. This is further elaborated on in Section 4.5.

4.4. Feature Analysis

This section aims to provide an in-depth analysis of the most significant features for predicting procedure duration. Section 4.4.1 presents the characteristics of the acquisition frequency, whereas Section 4.4.2 looks at the longitudinal position of the operating table. Section 4.4.3 presents the feature that concerns cumulative procedure time, followed by patient age in Section 4.4.4 and finally cumulative cine time in Section 4.4.5. The interpretation of the overall results is presented in Section 4.5. Furthermore, Section 4.6 presents the author’s perspective on the clinical applicability of the model.

4.4.1. Acquisition Frequency

The acquisition frequency has proven to be the most important feature for analysing procedure duration, based on the current classification model. The most common values for acquisition frequency lie below 0.1 acquisitions per second. This observation is made based on the very concentrated area on the histogram in Figure 4.16a. The peaks in acquisition frequency at the values of 0.5 and 1 are probably from the first acquisition of the procedures; the point at which cumulative procedure time is still relatively low. The boxplot in Figure 4.16b shows that the acquisition frequency is lower for procedures of a higher duration. Furthermore, the interquartile range and the number of outliers decrease for procedures of a longer duration. This is especially visible for procedures that take longer than 30 minutes to complete. Additionally, the median values seem to be fairly equally distributed with respect to the interquartile range, for the highest three duration classes. For procedures under 10 minutes, the data is slightly skewed left, indicating a higher concentration of data points to the left of the median. This implies that the values of the lower acquisition frequencies are closer together than the slightly larger values, which can be confirmed by Figure 4.16a.

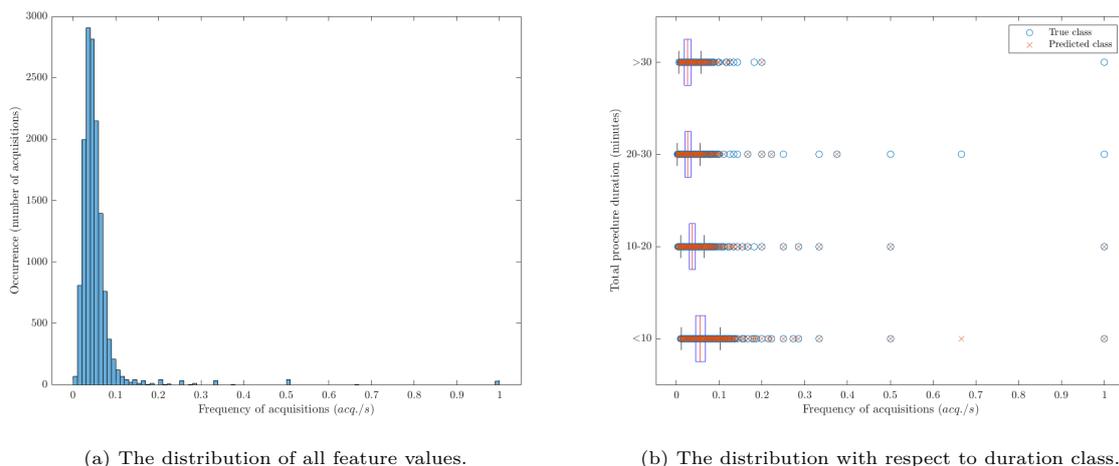


Figure 4.16: The distribution of all values for acquisition frequency within the dataset.

Figure 4.17 shows the general change in value for acquisition frequency throughout the procedure. Rather than plotting against cumulative procedure time, procedure progress is used, in order to take into account the varying duration among the procedures. As to achieve more insight into the variation among different duration classes, the classes are analysed individually. Average values for each class are calculated in segments of 5%, in terms of procedural progress. This means that twenty values are used for the generation of each curve. As observed from previous graphs, procedures with a shorter total duration have a higher acquisition frequency. Within the first 10% of the procedural progress, values for acquisition frequency show a steep decrease. This is because cumulative procedure time is relatively low, resulting in a high value for frequency. As time into the procedure goes by, fewer acquisitions are made with respect to the passed time of the procedure. Further on in the procedure, the values for acquisition frequency do not seem to change significantly. The curves that represent the first, third and fourth duration class seem to show slightly more variation throughout the procedure, whereas the acquisition frequency for procedures that last between 10 to 20 minutes seems more stable.

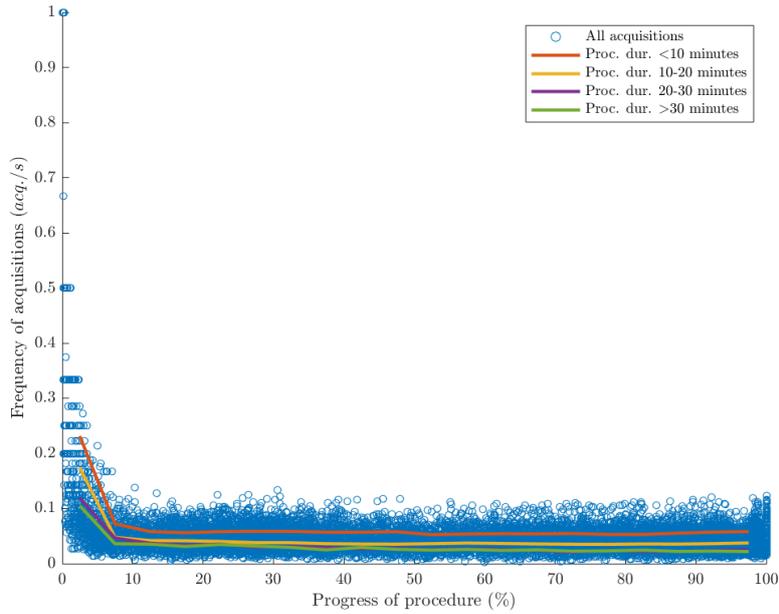
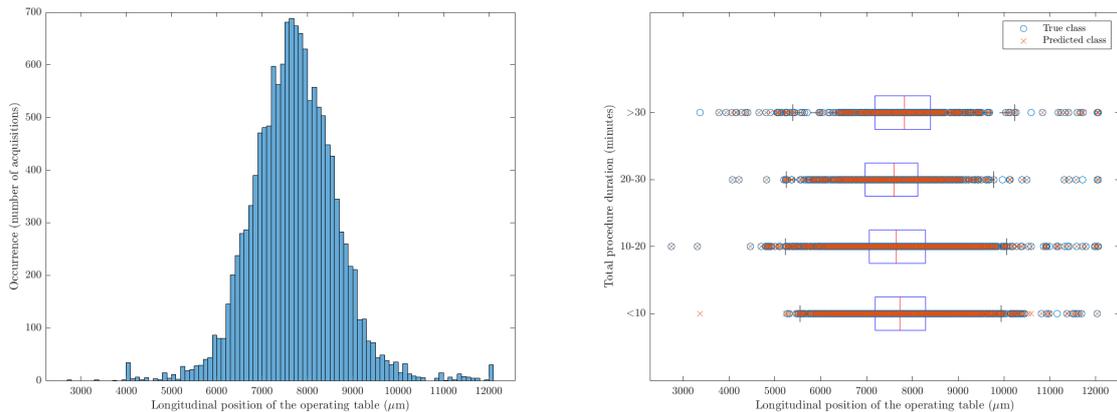


Figure 4.17: The values for acquisition frequency with respect to progress and duration class.

4.4.2. Longitudinal Position of the Operating Table

The second feature to be analysed includes the longitudinal position of the operating table. The histogram in Figure 4.18a shows the range of values for this feature, which seem to resemble a normal distribution. This implies that the degree to which the operating table moves towards and away from the interventional X-ray system, with respect to the mean position of the table, is roughly equal for both directions. This is also shown by the fact that the interquartile range and whiskers are approximately of the same length, on both sides of the median represented by the red vertical line. For all duration classes, the median value for table position lies between 7500 and 8000 μm . The median is shifted towards lower values of longitudinal table position for procedures that take between 10 and 30 minutes. The median position for procedures under 10 minutes is slightly higher, whereas the median for procedures over 30 minutes is the highest.



(a) The distribution of all feature values.

(b) The distribution with respect to duration class.

Figure 4.18: The distribution of all values for longitudinal table position within the dataset.

It is interesting to note that the changes in position of the operating table are relatively small. The total range of recorded values is 9 mm, which is minimal compared to the size of the operating table and the movements made by the X-ray system. Nevertheless, the small changes in position of the operating table seem to have powerful predictive abilities.

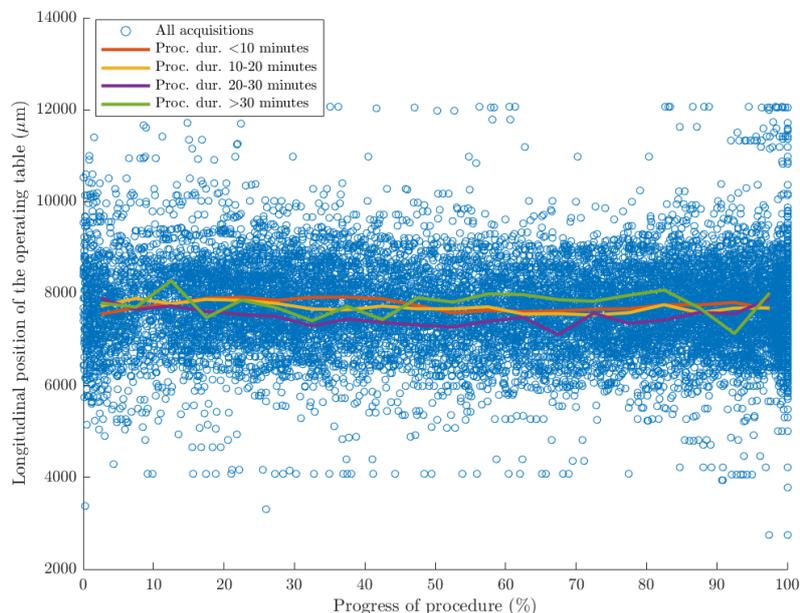


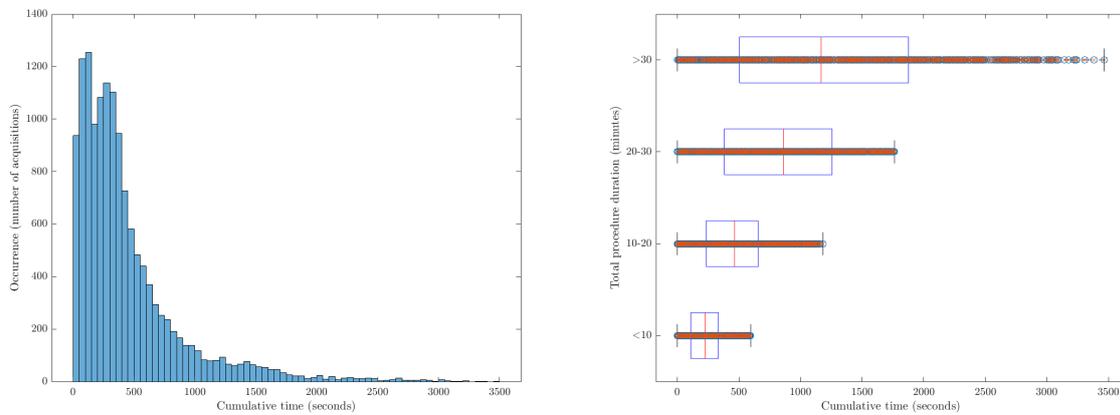
Figure 4.19: The values for longitudinal table position with respect to progress and duration class.

Figure 4.19 shows the general longitudinal position of the table with respect to progress of the procedure. It is noticeable that curves representing a procedure duration between 0 to 20 minutes show a low degree of variation throughout the procedure. Generally, the lateral position slightly increases in the first 50% of the short procedures, after which it slightly increases to end up at roughly the starting point. Procedures with a duration of 20 minutes or more show much more variation throughout the procedure. The curve that corresponds to the third duration class shows a relatively big decrease in terms of table position, followed by significant variability between 60% and 80% of the progress. Nevertheless, this variability could be due to one or two outliers, rather than the general pattern. Procedures with a duration of 30 minutes or over show the biggest variation. Again, this could be due to extreme outliers, or due to the fact that the table is repositioned more often during procedures that take longer. The boxplot in Figure 4.18b does not show a significant abundance and larger extremity of outliers for procedures that take over 30 minutes, with respect to the other duration classes. Therefore, it is plausible that procedures with a higher duration generally involve more longitudinal table movement.

4.4.3. Cumulative Procedure Duration

Cumulative procedure duration is a feature that is relatively simple to log, but one that is not part of the current data from the interventional X-ray system. Only the duration of each individual acquisition is logged, as of yet. Unsurprisingly, cumulative procedure duration is a useful feature for predicting the total duration. Figure 4.20a shows the distribution of logged cumulative procedure times. Every acquisition is processed and taken into account, so this histogram contains multiple data points from the same procedure. All procedures contain acquisitions at the start, which explains the high prevalence on the left side of the histogram. The occurrence decreases as the number of procedures with a cumulative duration that corresponds to each segment decreases.

Needless to say, the boxplot in Figure 4.20b show that the values for cumulative duration increase as the total procedure duration increases. The interquartile range and presence of outliers also increase with cumulative duration.



(a) The distribution of all feature values. (b) The distribution with respect to duration class.

Figure 4.20: The distribution of all values for longitudinal table position within the dataset.

Figure 4.21 shows how cumulative procedure duration changes throughout the procedure, with respect to duration class. Among the duration classes, the higher classes seem to show more variation throughout the procedure. The curves that correspond to procedures between 0 to 10 and 10 to 20 minutes are relatively straight. The variability within the two highest classes is likely to be due to the fact that these classes have more outliers.

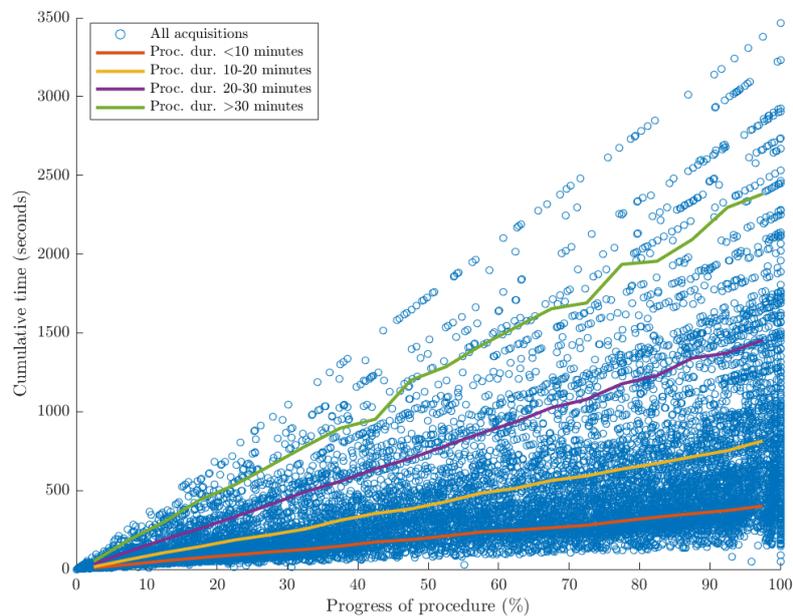


Figure 4.21: The values for cumulative procedure duration with respect to progress and duration class.

4.4.4. Patient Age Range

Since patient age does not vary among acquisitions of the same patient, this feature is plotted with respect to procedure rather than each acquisition. Furthermore, the third graph is not included, for the same reason.

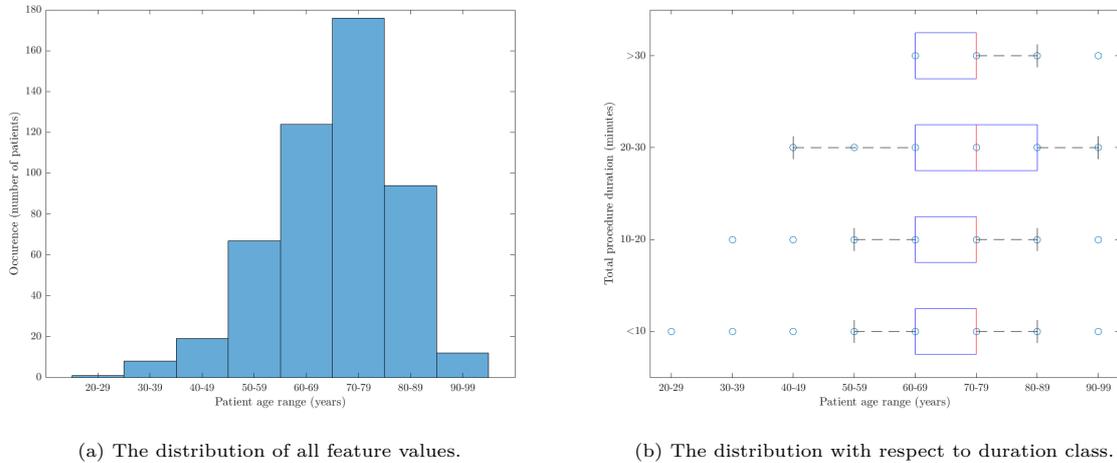
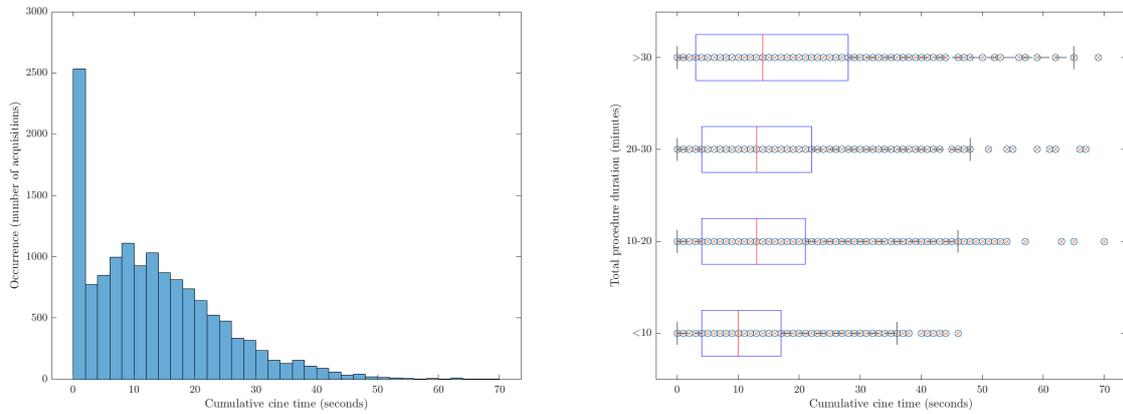


Figure 4.22: The distribution of the age of all patients within the dataset.

Figure 4.22a shows that most procedures involve patients between 70 and 79 years old. The second most prevalent age group represents patients of 60-69 years old, followed by patients of age of 80-89. Procedures with patients under 40 or 90 and over are relatively uncommon. According to Figure 4.22b, all procedures done on patients younger than 40 years old were completed within 20 minutes. All procedures that took longer than 30 minutes involve patients of 60 and older. The median age group of all procedure duration classes is the 70 to 79 group. The third duration class shows a distribution in values that is slightly less concentrated than the other three. Data points in the highest duration class, of 30 minutes and over, are the most concentrated. The observations show an undeniable correlation between patient age and total procedure duration, which explains the significance of the feature for predicting procedure duration.

4.4.5. Cumulative Cine Acquisition Time

A histogram of the cumulative duration during which the cine application is used, is presented in Figure 4.23a. The cumulative cine times range from 0 to 70 seconds. Comparing the values for cumulative cine time of duration class 1 to duration class 4, the median value for cumulative cine time increases from approximately 10 seconds to 16 seconds. This can be seen in Figure 4.23a. Furthermore, the interquartile range and range of outliers increase as the duration class increases as well.



(a) The distribution of all feature values.

(b) The distribution with respect to duration class.

Figure 4.23: The distribution of all values for patient age range within the dataset.

Figure 4.24 presents the general values for cumulative cine time with respect to procedure progress. As was the case for the general cumulative procedure time graph, the higher duration classes show more variation throughout the procedure. Especially the curve that corresponds to Class 3 procedures shows significant instability. This can again be explained by the fact that the data points for procedures of over 30 minutes are more widespread and therefore contain more outliers.

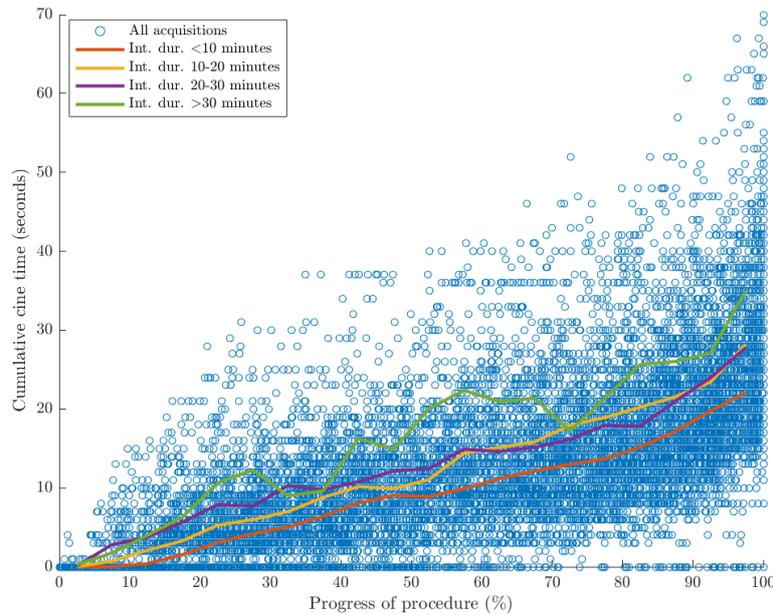


Figure 4.24: The values for cumulative procedure duration with respect to progress and duration class.

4.5. Interpretation of the Feature Analysis

The analyses from Section 4.4 help to understand how the most important features change with respect to procedure duration. Furthermore, they provide numerical representations of the features and how these vary per duration class. As for acquisition frequency, the most significant feature for the model, values are higher for procedures of a shorter duration. Furthermore, the acquisition frequency shows a peak at the start of the procedure and remains stable throughout the rest of it. For some of the features within the dataset, the change with respect to duration is less apparent than for others. An example is the longitudinal table position, which shows no clear patterns with respect to duration class. Nevertheless, the random forest model considers it to be the second most important feature. This shows that a lot of information can be extracted from the model, but some of its black-box characteristics remain. For cumulative procedure time and cumulative cine time, the general patterns are more straight-forward. Their respective graphs are self-explanatory and show clear differences with respect to each duration class. The final feature that is analysed in further depth is patient age. Most procedures have been performed on patients between the age of 60 and 90. None of the extreme delays have occurred in procedures that involved patients below 60 years of age. Nevertheless, procedures that took under 10 minutes have been performed on all ages. Further research into the effect of patient age could be used for preoperative scheduling, where the expected procedure duration is personalised based on patient characteristics. This is something that is currently not done at the Reinier de Graaf hospital, as the approximate procedure duration of a CAG is estimated at 45 minutes for all patients.

In order to take the use and analysis of individual features to a new level, pattern recognition could be applied to find new correlations between features and procedure duration or workflow. For example, phases could be defined in terms of other features besides angulation and rotation, as has been done in Section 3.1.5. Based on the RFE analysis, neither feature was included in the final feature set, which means that they were not accurate enough to be considered useful for the classification model. By analysing progress in terms of other features than time, individual sources of delay could be detected that could, apart from just stating a delay, be used to explain why a particular delay occurs. This is not an attribute of the current model, but could be of significant value for workflow analysis.

Medical interpretation

The results that have been found in this research project have been discussed with the cardiology department of the Reinier de Graaf hospital, in order to gain a better understanding of the reason that features change throughout a procedure. An interesting finding in this report includes the fact that the operating table provides relevant information about the procedural progress. As a matter of fact, the analysis states that the longitudinal table position is the second most important feature for predicting total duration, whereas features that regard the position of the interventional X-ray machine are not included in the five most important features. According to cardiologist Dr. Gilly, the main reason for moving the operating table is to place the tip of the catheter at a particular point with respect to the entire picture frame on the monitor. This ensures that once the contrast fluid is injected, all desired arteries will be displayed on the monitor. The correct display position is achieved by first repositioning the interventional X-ray system, including the C-arm, followed by a change in position of the operating table. Since a standard CAG requires multiple views of each coronary artery, the table position is changed almost every time. All three directions of the operating table are moved independently. The operating table is predominantly moved in the longitudinal and lateral directions and generally requires less vertical displacement. Furthermore, the cardiologists often require a bit of time to adapt to the position of the equipment combined with the external and internal characteristics of the patient. Therefore, the start of a procedure often shows more movement of the equipment. The same can be said about the start of a new phase. Generally, the cardiologists start off by capturing the RCA, which requires some adaptation but contains fewer steps and acquisitions. Once the LCA phase commences, the table again requires some extra adjustments to ensure that the patient as well as the equipment are in the right position. Since the LCA phase requires more acquisitions, the table is likely to show more movement in this part of the procedure. However, Figure 4.19 does not show more fluctuations in the second half. What can be observed is that the table seems to fluctuate more for procedures of a higher duration. According to Dr. Gilly, this could indeed be because more movement is required to find the optimal position for adequate display, which leads to a higher procedure duration.

With regard to patient age, Dr. Gilly confirms that for older patients, vessels are brittle and pliable and contain more turns. This can complicate the insertion of the guidewire and the catheter. Furthermore, the presence of plaque can make it more difficult to enter the vessels. Nevertheless, the aforementioned situations are only the case for some patients of age and the effect is not generalisable for any specific age groups. Furthermore, access to the vessels might be hindered, but the CAG does not necessarily require more acquisitions with respect to a younger patient. However, the fact that a higher age induces a higher risk of obstructive situations is acknowledged. Furthermore, the width of the aorta increases with age, which can lead to undesired withdrawal of the catheter. In all likelihood, these are the reasons for the significant effect of patient age on predicted duration. Another patient feature that has shown to be an important feature for determining duration, is object thickness. According to medical experts Dr. Gilly and Dr. Hoftijzer, patients that lead the X-ray system to detect a higher object thickness are more difficult to capture. This is because more radiation is required to penetrate the tissue and obtain fluoroscopic recordings of good quality. Often, this implies that more acquisitions are required, leading to a longer procedure duration.

Alternative sources of information

Section 3.1.4 elaborates on features that could be added to the dataset to provide more insight into procedural workflow. This is information that can be calculated based on the data logged by the interventional X-ray system. Nevertheless, there are further sources of information in the cath lab that could be used to obtain information regarding the procedure. First of all, images recorded by the interventional X-ray system could provide relevant information regarding procedural progress. For example, image recognition could be applied to detect a particular artery on the image, which can then be linked to a procedural step of the CAG. This could result in precise evaluation of interventional workflow. If the recordings were to be obtained and analysed real-time, it could even be used for intraoperative duration prediction. Another viable source of information is the operating table. Currently, only positional parameters are tracked. Nevertheless, it could provide more information regarding the patient, such as patient weight. The data logged by the interventional X-ray system does contain a column for patient weight but the content is empty, implying that patient weight is not yet included. If a scale were to be implemented into the operating table, it could also be used to detect when the patient lies down and when the patient exits the room. This would give a better indication of the total procedure duration than acquisition times or the time at which the patient file is uploaded on DoseWise.

The overarching workflow project, discussed in Section 2.2, aims to use camera footage to analyse procedural workflow and cath lab efficiency. This provides information on the presence and movements of patients and medical staff. Furthermore, the footage could be used to detect specific steps of the procedure, based on hand movement or other actions taken by the physicians. Currently, the cameras only record the inside of the cath lab. This means that there is no registration of movement inside the control room or in the patient waiting area. Further observation in these areas would provide additional insight into the throughput of patients and cath lab efficiency. Alternatively, laser sensors could be used to detect movement. This would result in data that is less informative and detailed, yet it would lead to fewer issues with respect to patient privacy. Rather than tracking the actors inside the room, one could also aim to observe other elements. For example, medical instruments could be tracked by means of radio-frequency identification or other techniques. The use of instruments or tools can be linked to the interventional step in which they are commonly used, providing new information regarding procedural workflow.

With respect to the data logged by the interventional X-ray system, there are several elements that would further increase the usefulness of the dataset for workflow analysis and duration prediction. First of all, the additional features from Section 3.1.4 add valuable information regarding previous acquisitions of the same patient and their implementation is recommended. Normally, a random forest classification algorithm does not take this information into account, whereas it improves the predictions significantly. Furthermore, it would be useful to log procedure type, to allow for real-time analysis without having to confirm the procedure type using DoseWise software. Implementing this into the input data for the random forest algorithm would lead to more distinction between different procedures and a better model performance.

More specifically, information on the type of CAG, such as CAG femoralis or bypass CAG, would benefit the dataset even further. According to the cardiology department, a planned CAG femoralis does not necessarily result in a longer procedure. Nevertheless, the procedure is often delayed when the physician first attempts a CAG radialis, to then divert to the groin. Bypass CAG procedures do usually take longer than the average duration discussed in this report. Finally, additional offline data could be added to the dataset used for the random forest algorithm. Examples include the composition of staff members, their experience and the time of the day. This data could be used to extend the feature set for training of the random forest model, as well as for preoperative scheduling.

4.6. Clinical Implementation of the Model

The random forest classification model that has been created and tested in this report, aims to accurately predict the total procedure duration based on intraoperative acquisition data. Based on the found results, the model is particularly useful for detecting delay, which refers to procedures with a duration of 20 minutes or more. Based on the used dataset, such delays are commonly detected within the first five minutes of the procedure. Furthermore, the model does not tend to overestimate procedure duration beyond a cumulative procedure time of 10 minutes. This implies that predictions made after this time provide the minimum duration class of the procedure. For example, when the procedure is estimated to take 20 to 30 minutes, the absolute minimum duration is 20 minutes. Nevertheless, the prediction might shift to a duration of 30 minutes or more, in which case the minimum duration of the procedure is now 30 minutes. In a clinical setting, this provides information regarding the minimum amount of time for which the cath lab will still be occupied. This can then be applied for dynamic scheduling of the cath lab and for knowing the minimum amount of time one should wait before calling the next patient.

For procedures that take less than 20 minutes, the model still shows a high prediction accuracy. As a matter of fact, predictions for Class 1 and Class 2 procedures are more accurate than those for delayed procedures, with a prediction accuracy of 98.1% and 89.8%. Nevertheless, results have shown that within the first ten minutes of the procedure, 1.9% of the data points from Class 1 are overestimated. Therefore, the assumption that the predicted values provide the minimum procedure duration is not reliable within this time frame. Within the dataset, 61.6% of all procedures take less than 10 minutes. Once the cumulative procedure time has exceeded the 10-minute margin, Figure 4.4 shows that Class 2 acquisitions are no longer overestimated, but the usefulness of the predictions at that point are questionable. Once the procedure time exceeds 10 minutes, the physician does not require a model to understand that the minimum duration is more than 10 minutes. Nevertheless, the tool can be used to update third parties on the status of the procedure without requiring verbal communication, which could possibly distract the staff from the procedure itself.

Since the classification model is mainly useful for detecting and predicting delay, it is important to understand how often procedures are delayed before evaluating whether model implementation is valuable. Table 3.6 shows that for 11.52% of the procedures within the used dataset, the operative phase took longer than 20 minutes to be completed. With an average of 4.2 CAG procedures per day, based on the distribution shown in Appendix B, approximately 4 to 5 procedures are delayed by more 20 minutes per week. With respect to the overall occupation of the cath lab on a weekly basis, the model should undergo further optimisation before it can be assessed as a viable prediction tool. For clinical implementation, a model would need to be built for real-time extraction and analysis of the interventional X-ray data and user interface must be designed to transfer the information to the physicians. It is currently uncertain as to whether the efficiency increase is big enough to cover the costs of model implementation. In terms of model optimisation, further research must be done into the additional value of the model, with respect to the physician's own interpretation of the situation. Furthermore, the accuracy of predictions made within the first 10 minutes of the procedure should be improved, to make the model suitable for procedures of a normal duration as well.

In terms of the advantages of the model to the hospital, the main one is that it can predict procedural duration and delay. This allows for the creation of a dynamic schedule rather than a static one, requiring less manual scheduling work due to unforeseen circumstances. Furthermore, the model can be set to keep learning from new procedures, which means that data is actively used to continuously improve duration predictions. Currently, the model is mostly suitable for indicating progress and predicting delays of 20 minutes or above. Nevertheless, a perfect model would allow for accurate prediction for the duration of any procedure. It could notify the physicians and schedulers about predicted duration as well as delays, at an appropriate amount of time in advance. Of course, it can never be completely robust to last-minute changes or complications. Nevertheless, the logged data should be able to provide enough information to generate an approximate value for procedural duration.

As far as the generalisability of the model is concerned, it is currently only trained on data from CAG procedures. For the application of the algorithm to CAG procedures from different time periods or different hospitals, the model is assumed to be applicable and sufficiently generalised. It is trained on a wide variety of CAG data and hyperparameter tuning has shown that overfitting is reduced to a minimum. Nevertheless, an additional system is required that calculates the additional features as part of the input data, as these features are not currently part of the overall dataset. For intraoperative prediction of procedure duration, these features must be calculated and extracted real-time. With regards to different procedure types executed inside in the cath lab, the data format is identical. Therefore, if the specific procedures were filtered from the data and the entire process described in this report would be repeated, the resulting model could be applied to predict the duration for other procedure types. However, it is important that the process of training the model is repeated and that the parameters and optimal feature set are reevaluated using the new dataset. The final model created in this report, including the hyperparameters and the feature set based on recursive feature elimination, is created and optimised for CAG data only. Finally, the application of the model to other procedures that occur outside the cath lab, such as surgeries in the operating department, requires further alteration of the model. First of all, the data is likely to originate from other equipment than the C-arm. Therefore, its format must be evaluated before an optimisation of the feature set can occur. For the interventional X-ray data, several features had to be converted to allow for their implementation into the random forest algorithm in Matlab. Furthermore, specific procedures would be required that generate multiple data points per patient, in order to be linked to operative workflow. In conclusion, the applicability of the model cannot be considered universal, as it depends on the type of input data and the procedure type. Nevertheless, the steps for data collection and model optimisation are the same. Therefore, a large part of the Matlab code could be reused.

5. Conclusion & Recommendations

Section 5.1 presents the conclusions of this report and answers the research question stated in Chapter 1. Section 5.2 presents the recommendations for future research within this project.

5.1. Conclusion

The main aim of the research presented in this report was to analyse the usability of interventional X-ray data for intraoperatively predicting procedure duration of a CAG. Furthermore, the goal was to analyse the most important features in further depth, as to understand how they change with respect to progress and different duration classes. The available dataset from the interventional X-ray system contains 601 CAG procedures, characterised by 16,536 acquisitions and 30 useful features. A total of 22 additional features were generated that implemented the information from previous acquisitions of the same procedure, referred to as repeated measures. Subsequently, the optimal combination of features was deduced using recursive feature elimination, resulting in a final feature set of 19 features.

In order to answer the research question: *'To what extent can data logged by interventional X-ray systems be used to intraoperatively predict the total procedure duration class of a CAG and how do the most significant features affect the analysis?'*, a random forest classification model was used. For each data point in the interventional X-ray data, the model predicted whether the procedure would take 0 to 10 minutes, 10 to 20 minutes, 20 to 30 minutes or more than 30 minutes. Based on out-of-bag validation, the model showed an overall classification accuracy of 92.75%. Given the fact that the interventional X-ray data is mainly logged for machine maintenance and repair, this is exceptionally high. The model has some difficulties in correctly classifying procedures that take under 10 minutes, where it predicts a procedure duration of 10 to 20 minutes in 1.9% of the cases. For the remaining three classes, the model only overestimates procedure time in less than 0.5% of the cases. Therefore, once a duration class has been estimated beyond the 10-minute margin, the estimation can be considered reliable. Furthermore, Class 3 and Class 4 predictions provide a reliable indication of minimum procedure duration from the start. Nevertheless, the average operative duration of a CAG is 11 minutes. Therefore, the model seems more applicable to delayed procedures, rather than to procedures of a standard duration.

For further analysis of model performance, the intraoperative predictions have been analysed with respect to cumulative procedure time. Incorrect classification of Class 1 acquisitions seems to occur until the 10-minute margin. For the other 98.1% of the acquisitions from procedures under 10 minutes, the correct class is generally predicted within the first 2 minutes. For the observations from procedures that took between 10 and 30 minutes, misclassifications occur until the start of the correct class, thus 10 and 20 minutes respectively. Nevertheless, the majority of the predictions for Class 2 and Class 3 are made before a cumulative procedure duration of 5 minutes. Despite a small degree of uncertainty, the model shows a high accuracy when predicting for Class 2 and Class 3. For procedures that take over 30 minutes, misclassifications occur until the 20-minute margin. Nevertheless, the majority of the Class 4 predictions occur within the first 6 minutes of a procedure. Predictions that state a duration of 30 minutes and over have an accuracy rate of 99.9%. In terms of predictive abilities, predictions for extremely delayed procedures are therefore very accurate and definitely competent for practical application.

When analysing individual procedures, it seems that the model's applicability varies per procedure. Each analysis is done by a different random forest classification model, that is trained using all available data except for the acquisitions from the analysed procedure. One procedure shows incorrect predictions that were not made in previous analyses. A possible reason is all acquisitions from the procedure were excluded from the training set, removing any possible optimistic bias. It could explain why some of the individual results seem slightly less useful with respect to the general results. This would imply that acquisitions from the same patient are too dependent to be split into training and test data. Further research would have to show whether the repeated measures do indeed bias the results, or whether the individual analyses are different based on an alternative reason.

Within the analysis of individual procedures, some intraoperative predictions for procedures that take under 10 minutes shift from Class 1 to Class 2 and vice versa. Due to this fluctuating output, the application of the model is not as useful for shorter procedures. Nevertheless, for other individual procedures, the model is able to estimate delay from a relatively early point in the procedure. Additionally, for some procedures that take over 10 minutes, the initial observations are immediately classified as Class 2 or over. Since the model generally does not overestimate after a cumulative procedure time of 10 minutes, the first prediction of a new class beyond this time can be considered a reliable reference. From this point onwards, it is safe to assume that the new minimum duration of the procedure lies within the range of this newly predicted class.

In terms of feature importance for estimating duration class, the acquisition frequency was found to be the most important feature, followed by the longitudinal position of the operating table and cumulative procedure duration. Patient age and cumulative cine acquisition duration time were also found to be significant. Among these features, the acquisition frequency and cumulative duration values show a clear pattern throughout each procedure. Furthermore, their difference with respect to procedures of varying duration classes is evident. Nevertheless, the effect of table position on the prediction is not directly apparent. Further analysis into these findings, that are partially hidden behind the black box characteristics of the model, would have to be done for a full understanding of the feature's contribution.

To answer the research question, data from the interventional X-ray system is definitely useful for predicting progress and the total duration of a CAG procedure. For procedures that take less than 10 minutes, the random forest model has difficulties in accurately predicting procedure duration. As most procedures have an operative length of less than 10 minutes, more research must be done into optimising predictions within this duration range. This optimisation should be implemented before the model can be considered viable. Nevertheless, the classification model shows exceptionally predictive properties for determining delay. Most delays are detected and correctly classified within the first five minutes of the procedure. These characteristics could be extremely useful for an efficient throughput of the cath lab and could be further developed to create a tool for analysis and prediction of CAG workflow. The original features logged by the interventional X-ray system are informative and show promising abilities when it comes to intraoperatively classifying duration. Furthermore, the addition of features that take into account repeated measures have shown to improve the classification accuracy significantly. As the model is able to keep track of procedures progress, the model could also be used for updating third parties on the workflow of the procedure, such as hospital schedules. This would eliminate the need of personal communication with the cath lab, in which the staff can now maintain their focus on the procedure itself. The main area of improvement of the classification model concerns the prediction of duration of procedures that are not delayed. For future research on this or on the usability of interventional X-ray data in general, this report provides a useful and informative starting point.

5.2. Recommendations

This section provides recommendations for future research, that exceeded the scope of this research. One element that could be further examined is the use of other machine learning methods that are particularly suitable for datasets containing repeated measures, such as (non-)linear regression. Also, the random forest model could be tested using a smaller class size, to test the extent to which the model can predict duration based on shorter time intervals. Using smaller classes could be particularly beneficial for analysing short procedures. Another improvement could be to link more features to progress or duration. Currently, the phases of a CAG have only been defined in terms of the angulation and rotation of the C-arm. Nevertheless, the operating table has shown to contain an equal number of informative properties as the X-ray system. Pattern recognition could be applied to define phases with respect to such features. These phases could then be implemented to analyse workflow, or to improve the input data of the current random forest model. This could possibly increase the ability of the model to detect a delay or predict duration earlier on in the procedure. Further research into this subject would provide more information on the feasibility of predicting duration early on in the procedure. The minimum time at which workflow can be analysed and unforeseen events can be predicted is currently unknown.

In terms of the input data, relevant information could be added by indicating CAG procedures in which the physician had to divert to the groin or where the arteries were accessed through a bypass. By implementing this information into the input data, the model could be trained to detect a delay due to the diversion to another type of CAG. Alternatively, the effect of the repeated measures in the data could be further investigated, including the similarities between acquisitions from the same patient. One could also experiment with datasets from other equipment, such as medical instruments. With respect to the CAG data, it is recommended to include data from previous acquisitions within the procedure when using a random forest model. The additional features are not expected to increase the degree of bias from repeated measures. Furthermore, the system should record the time at which the patient enters the room or when the electronic patient record is uploaded. This would provide relevant information regarding the occupancy of the cath lab. Also, procedure type should be implemented into the interventional X-ray data, to avoid the need of validation using DoseWise software. Alternatively, research could be done into other methods to determine procedure type.

To expand the understanding of the classification model, more testing could be done on new datasets using cross-validation, in order to fully remove the influence of repeated measures. Furthermore, the model could be trained and tested using data from different procedures that also consist of multiple acquisitions per patient. Moreover, testing could be done on CAG procedures from different hospitals. The results could provide hospital-specific information regarding feature importance and workflow. For a better insight into the functionality of the current model, available hospital schedules could be used to compare the original planning to the outcome, with and without the applied classification model. Furthermore, the effect of including specific procedures could be analysed. For example, procedures that contain a few acquisitions would train the model differently than procedures with many acquisitions. The effect of such differences might influence the classification accuracy. Overall, it is favourable to include all data types in the training data, as to represent a realistic dataset as closely as possible.

To take the current classification model up a notch, a random forest regression model could be developed and tested. Rather than classifying delay, this would predict a numerical value for total procedure duration. Nevertheless, attempts have shown that more information regarding the data and its phases would be required to obtain accurate predictions. Alternatively, the classification model could be combined with information regarding the workflow of a CAG in terms of other parameters. Currently, progress is only measured with respect to time. Using workflow analysis and the aforementioned pattern recognition tool to analyse the information from alternative features, a combined model could be created for detailed workflow analysis, rather than a tool for duration prediction. To conclude, results regarding the usability of the interventional X-ray data are highly promising and further research into the application of the data is recommended.

Bibliography

- [1] Matthew A. Bartek, Rajeev C. Saxena, Stuart Solomon, Christine T. Fong, Lakshmana D. Behara, Ravitheja Venigandla, Kalyani Velagapudi, John D. Lang, and Bala G. Nair. Improving Operating Room Efficiency: Machine Learning Approach to Predict Case-Time Duration. *Journal of the American College of Surgeons*, 229(4):346–354.e3, 2019. ISSN 1879-1190. doi: 10.1016/j.jamcollsurg.2019.05.029.
- [2] Raymond R. Bond and Aaron Peace. Near future artificial intelligence in interventional cardiology: new opportunities and challenges to improve the care of STEMI patients.
- [3] L. Breiman. Out-of-Bag Estimation. Technical report, University of California, Berkely, CA, 1996. URL [/paper/OUT-OF-BAG-ESTIMATION-Breiman/e408af07b74476564e3d5511ed6b169fa1f2a484](#).
- [4] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, October 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. URL <https://doi.org/10.1023/A:1010933404324>.
- [5] Emmanouil Brilakis. *Manual of Chronic Total Occlusion Interventions: A Step-by-Step Approach*. Academic Press, November 2017. ISBN 978-0-12-809930-8. Google-Books-ID: LEyZDgAAQBAJ.
- [6] Jason Brownlee. How to Reduce Variance in a Final Machine Learning Model, August 2018. URL <https://machinelearningmastery.com/how-to-reduce-model-variance/>.
- [7] Ping-Shun Chen, Ying-Jie Lin, and Nai-Chun Peng. A two-stage method to determine the allocation and scheduling of medical staff in uncertain environments. *Computers & Industrial Engineering*, 99:174–188, September 2016. ISSN 0360-8352. doi: 10.1016/j.cie.2016.07.018. URL <http://www.sciencedirect.com/science/article/pii/S0360835216302534>.
- [8] Davuluri Hemanth Chowdary. Decision Trees Explained With a Practical Example Towards AI The Best of Tech, Science, and Engineering, May 2020. URL <https://towardsai.net/p/programming/decision-trees-explained-with-a-practical-example-fe47872d3b53>, <https://towardsai.net/p/programming/decision-trees-explained-with-a-practical-example-fe47872d3b53>.
- [9] Antonio Criminisi, Jamie Shotton, and Ender Konukoglu. Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. 7, January 2012.
- [10] Daniel Walter. Overfitting, Regularization, and Hyperparameters, November 2015. URL <https://dswalter.github.io/overfitting-regularization-hyperparameters.html>.
- [11] Burcu F. Darst, Kristen C. Malecki, and Corinne D. Engelman. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genetics*, 19(1): 65, September 2018. ISSN 1471-2156. doi: 10.1186/s12863-018-0633-8. URL <https://doi.org/10.1186/s12863-018-0633-8>.
- [12] Eberhard Zeitler and Ernst Ammann. *Radiology of Peripheral Vascular Diseases: With 198 Tables*. Springer Science & Business Media, 2000.
- [13] Marinus J. C. Eijkemans, Mark van Houdenhoven, Tien Nguyen, Eric Boersma, Ewout W. Steyerberg, and Geert Kazemier. Predicting the unpredictable: a new prediction model for operating room times using individual characteristics and the surgeon’s estimate. *Anesthesiology*, 112(1): 41–49, January 2010. ISSN 1528-1175. doi: 10.1097/ALN.0b013e3181c294c2.
- [14] Eric W. Fox, Ryan A. Hill, Scott G. Leibowitz, Anthony R. Olsen, Darren J. Thornbrugh, and Marc H. Weber. Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. *Environmental monitoring and assessment*, 189(7):316, July 2017. ISSN 0167-6369. doi: 10.1007/s10661-017-6025-0. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6049094/>.

- [15] Carol L. Goldstein, Michael Racz, and Edward L. Hannan. Impact of cardiac catheterization-percutaneous coronary intervention timing on in-hospital mortality. *American Heart Journal*, 144(4):561–567, October 2002. ISSN 0002-8703. doi: 10.1067/mhj.2002.125322. URL <http://www.sciencedirect.com/science/article/pii/S0002870302001230>.
- [16] Brandon Greenwell and Bradley Boehmke. *Random Forests / Hands-On Machine Learning with R*. Chapman and Hall, November 2019. ISBN 978-1-138-49568-5. URL <https://bradleyboehmke.github.io/HOML/random-forest.html>.
- [17] Annetje C. P. Guédon, M. Paalvast, F. Meeuwssen, D. Tax, A. Dijke, L. Wauben, M. V. D. Elst, J. Dankelman, and J. V. D. Dobbelsteen. It is Time to Prepare the Next patient Real-Time Prediction of Procedure Duration in Laparoscopic Cholecystectomies. *Journal of Medical Systems*, 2016. doi: 10.1007/s10916-016-0631-1.
- [18] Sunwoo Han and Hyunjoong Kim. On the Optimal Size of Candidate Feature Set in Random forest. *Applied Sciences*, 9(5):898, January 2019. doi: 10.3390/app9050898. URL <https://www.mdpi.com/2076-3417/9/5/898>. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.
- [19] Seetharaman Hariharan and Deryk Chen. Costs and Utilization of Operating Rooms in a Public Hospital in Trinidad, West Indies. *The Permanente Journal*, 19(4):e128–e132, 2015. ISSN 1552-5767. doi: 10.7812/TPP/14-183. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5293130/>.
- [20] Peter Hertrich. *Practical Radiography*. John Wiley & Sons, July 2005. ISBN 978-3-89578-210-7. Google-Books-ID: GSd0IqSt3bsC.
- [21] Jake Hoarse. How is Splitting Decided for Decision Trees?, August 2018. URL <https://www.displayr.com/how-is-splitting-decided-for-decision-trees/>.
- [22] IBM Cloud Education. Machine Learning, July 2020. URL <https://www.ibm.com/cloud/learn/machine-learning>.
- [23] Sietske Inming. Staff and intervention scheduling for operating rooms with variable input parameters based on scheduling phase, objective and method. Technical report, Delft University of Technology, The Hague, April 2020.
- [24] Jaimie Bird and Arianna Forneris. Fluoroscopy, Digital Subtraction Angiography and Mammography. URL http://199.116.233.101/index.php/Fluoroscopy,_Digital_Subtraction_Angiography_and_Mammography.
- [25] Silke Janitza and Roman Hornung. On the overestimation of random forests out-of-bag error. *PLoS ONE*, 13(8), August 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0201904. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6078316/>.
- [26] Jennifer Papin. A Suggested Approach for Improving Flow in the Cardiac Catheterization Laboratory. *Cath Lab Digest*, 21(7), July 2013. URL <https://www.cathlabdigest.com/articles/Suggested-Approach-Improving-Flow-Cardiac-Catheterization-Laboratory>.
- [27] Ri-ichiro Kakiyama. A Noninvasive Alternative to Coronary Angiography: Myocardial Contrast Echocardiography Following Strain Map as a Gate Way to Myocardial Contrast Echocardiography Map. *What Should We Know About Prevented, Diagnostic, and Interventional Therapy in Coronary Artery Disease*, March 2013. doi: 10.5772/54076. Publisher: IntechOpen.
- [28] K.M. van der Graaf. Formalization and Quantification of the Workflow in the Catheterization Laboratory. Technical report, Delft University of Technology, March 2020.
- [29] Richard L. Kobus, Ronald L. Skaggs, Michael Bobrow, Julia Thomas, Thomas M. Payette, and Stephen A. Klimont. *Building Type Basics for Healthcare Facilities*. John Wiley & Sons, April 2008. ISBN 978-0-470-13541-9. Google-Books-ID: TBXMA8ndUXoC.
- [30] Krishni. A Beginners Guide to Random Forest Regression, June 2019. URL <https://medium.com/datadriveninvestor/random-forest-regression-9871bc9a25eb>.

- [31] Kristien M. van der Graaf. Formalization and quantification of the workflow in the catheterization laboratory. *Thesis for Delft University of Technology*, March 2020.
- [32] M. Bozkaya. *Business process analysis with semantic dotted chart*. PhD thesis, Eindhoven University of Technology, October 2011. URL <https://research.tue.nl/en/studentTheses/business-process-analysis-with-semantic-dotted-chart>.
- [33] Marsha L. Knapik. Cardiac Catheterization Laboratory Facility Design and Equipment Selection. *Cath Lab Digest*, 10(12), June 2008. URL <https://www.cathlabdigest.com/articles/Cardiac-Catheterization-Laboratory-Facility-Design-and-Equipment-Selection>.
- [34] C. Di Martinelly, P. Baptiste, and M. Y. Maknoon. An assessment of the integration of nurse timetable changes with operating room planning and scheduling. *International Journal of Production Research*, 52(24):7239–7250, December 2014. ISSN 0020-7543. doi: 10.1080/00207543.2014.916827. URL <https://doi.org/10.1080/00207543.2014.916827>.
- [35] MathWorks. Bag of decision trees - MATLAB - MathWorks Benelux, 2020. URL <https://nl.mathworks.com/help/stats/treebagger-class.html>.
- [36] Morton Kern. Angiographic Projections Made Simple: An Easy Guide to Understanding Oblique Views. *Cath Lab Digest*, 19(8), August 2011. URL <https://www.cathlabdigest.com/articles/Angiographic-Projections-Made-Simple-Easy-Guide-Understanding-Oblique-Views>.
- [37] N. Narayan. *A nearest neighbor based cold-deck imputation for X-ray tube wear estimation*. PhD thesis, Eindhoven University of Technology, November 2017. URL <https://research.tue.nl/en/studentTheses/a-nearest-neighbor-based-cold-deck-imputation-for-x-ray-tube-wear>.
- [38] Srihari S. Naidu, Herbert D. Aronow, Lyndon C. Box, Peter L. Duffy, Daniel M. Kolansky, Joel M. Kupfer, Faisal Latif, Suresh R. Mulukutla, Sunil V. Rao, Rajesh V. Swaminathan, and James C. Blankenship. SCAI expert consensus statement: 2016 best practices in the cardiac catheterization laboratory. *Catheterization and Cardiovascular Interventions*, 88(3):407–423, 2016. ISSN 1522-726X. doi: 10.1002/ccd.26551. URL <https://www.onlinelibrary.wiley.com/doi/abs/10.1002/ccd.26551>. [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ccd.26551](https://onlinelibrary.wiley.com/doi/pdf/10.1002/ccd.26551).
- [39] K. Okamoto, J. Ito, K. Sakai, and S. Yoshimura. The Principle of Digital Subtraction Angiography and Radiological Protection. *Interventional Neuroradiology*, 6(Suppl 1):25–31, November 2000. ISSN 1591-0199. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3685929/>.
- [40] Philips Healthcare. Allura Xper FD20 X-ray system, . URL <https://www.usa.philips.com/healthcare/product/HC722012CA/alluraxperfd20xrayssystem>.
- [41] Philips Healthcare. Dosewise | Philips Healthcare, . URL <https://www.usa.philips.com/healthcare/clinical-solutions/dosewise>.
- [42] Philips Healthcare. AlluraClarity Low-dose Interventional X-ray system, 2014. URL <https://www.philips.nl/healthcare/product/HCN0CTN163/alluraclarity-low-dose-interventional-x-ray-system/documentatie>.
- [43] Philips Healthcare. Instructions for Use of the Azurion, July 2017. URL https://philipsproductcontent.blob.core.windows.net/assets/Instruction%20for%20Use/20171107/fffdeacc907e4a9882d7a82400eeb113.pdf?feed=ifu_docs_feed.
- [44] Quizlet Inc. Diagram of Anterior and Posterior Coronary Arteries. URL <https://quizlet.com/295067074/anterior-and-posterior-coronary-arteries-diagram/>.
- [45] Dario Radei. Feature Selection in Python Recursive Feature Elimination, September 2019. URL <https://towardsdatascience.com/feature-selection-in-python-recursive-feature-elimination-19f1c39b8d15>.

- [46] Grant W. Reed, Scott Hantz, Rebecca Cunningham, Amar Krishnaswamy, Stephen G. Ellis, Umesh Khot, Joe Rak, and Samir R. Kapadia. Operational Efficiency and Productivity Improvement Initiatives in a Large Cardiac Catheterization Laboratory. *JACC: Cardiovascular Interventions*, 11(4):329–338, February 2018. ISSN 1936-8798. doi: 10.1016/j.jcin.2017.09.025. URL <http://www.sciencedirect.com/science/article/pii/S1936879817319957>.
- [47] Johan H. C. Reiber, P. W. Serruys, and C. J. Slager. *Quantitative Coronary and Left Ventricular Cineangiography: Methodology and Clinical Applications*. Springer Science & Business Media, December 2012. ISBN 978-94-009-4239-4. Google-Books-ID: 0ZiLBwAAQBAJ.
- [48] Binita Shah, Xingchen Mai, Lakshmi Tummala, Chad Kliger, Sripal Bangalore, Louis H. Miller, Steven P. Sedlis, Frederick Feit, Michael Liou, Michael Attubato, John Coppola, and James Slater. Effectiveness of Fluorography versus Cineangiography at Reducing Radiation Exposure During Diagnostic Coronary Angiography. *The American journal of cardiology*, 113(7):1093–1098, April 2014. ISSN 0002-9149. doi: 10.1016/j.amjcard.2013.12.013. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3959232/>.
- [49] Zahra Shahabi Kargar, Sankalp Khanna, Abdul Sattar, Norm Good, James Lind, and John O’Dwyer. Predicting Procedure Duration to Improve Scheduling of Elective Surgery. December 2014. doi: 10.1007/978-3-319-13560-1_86.
- [50] Sharma et al. Optimizing Catheterization Laboratory Throughput Using Machine Learning Patent Application. *USPTO.report*, August 2020. URL <https://uspto.report/patent/app/20200273569>.
- [51] David Shavelle. Basic Coronary Angiography. *Keck Medicine of USC*, page 55.
- [52] Fabio da Silva. *The role of R-spondin3 in coronary artery formation and novel roles for retinoic acid signaling in cardiac development and repair*. phdthesis, Université Côte d’Azur, October 2017. URL <https://tel.archives-ouvertes.fr/tel-01682211>.
- [53] Ralf Stauder, Asl Okur, Loïc Peter, Armin Schneider, Michael Kranzfelder, Hubertus Feussner, and Nassir Navab. Random Forests for Phase Detection in Surgical Workflow Analysis. In Danail Stoyanov, D. Louis Collins, Ichiro Sakuma, Purang Abolmaesumi, and Pierre Jannin, editors, *Information Processing in Computer-Assisted Interventions*, Lecture Notes in Computer Science, pages 148–157, Cham, 2014. Springer International Publishing. ISBN 978-3-319-07521-1. doi: 10.1007/978-3-319-07521-1_16.
- [54] Pieter S. Stepaniak, Christiaan Heij, and Guus De Vries. Modeling and prediction of surgical procedure times. *Statistica Neerlandica*, 64(1):1–18, 2010. ISSN 1467-9574. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9574.2009.00440.x>.
- [55] Sullivan Lisa M. Repeated Measures. *Circulation*, 117(9):1238–1243, March 2008. doi: 10.1161/CIRCULATIONAHA.107.654350. URL <https://www.ahajournals.org/doi/full/10.1161/circulationaha.107.654350>. Publisher: American Heart Association.
- [56] F. Van Luyn. Intra-operative estimation of surgical progress. 2017. URL <https://repository.tudelft.nl/islandora/object/uuid%3Ab418c732-fb2d-4a5f-a936-407a1ede45c7>.
- [57] Sandy Watson and Kenneth Gorski. *Invasive Cardiology: A Manual for Cath Lab Personnel*. Jones & Bartlett Learning, October 2010. ISBN 978-0-7637-6468-5.

A. Distribution of Acquisitions

The following graph shows the distribution of acquisitions with respect to procedural progress. The insight provides an additional perspective on the graphs that represent classification timing, as presented in Section 4.1.3. Figure A.1 shows that the number of acquisitions slightly increase between 20% and 40% of the procedural progress. After decreasing to a minimum between 40% and 50% of the procedure, the number of acquisitions increase exponentially, until a maximum is reached between 90% and 100% of the total procedure duration.

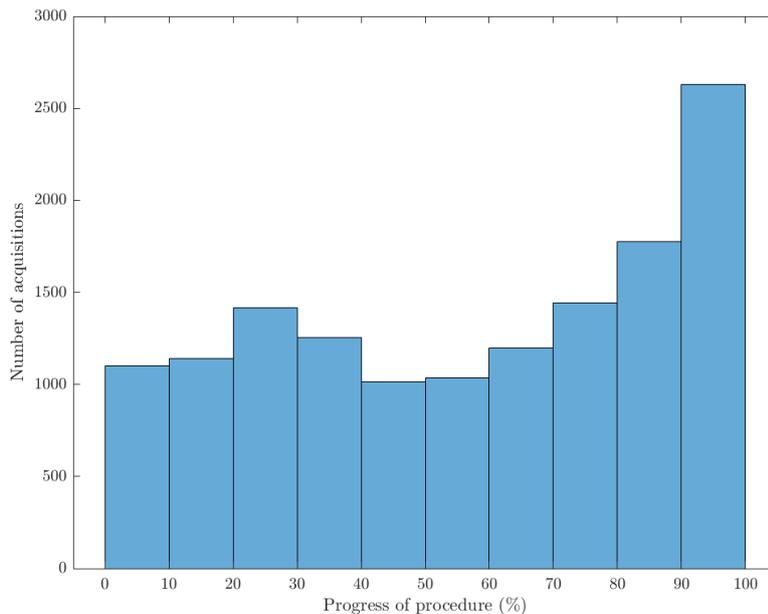


Figure A.1: The distribution of acquisitions with respect to procedural progress.

B. Distribution of Daily CAG Procedures

Figure B.1 shows a distribution of the number of CAG procedures per day, within the used dataset.

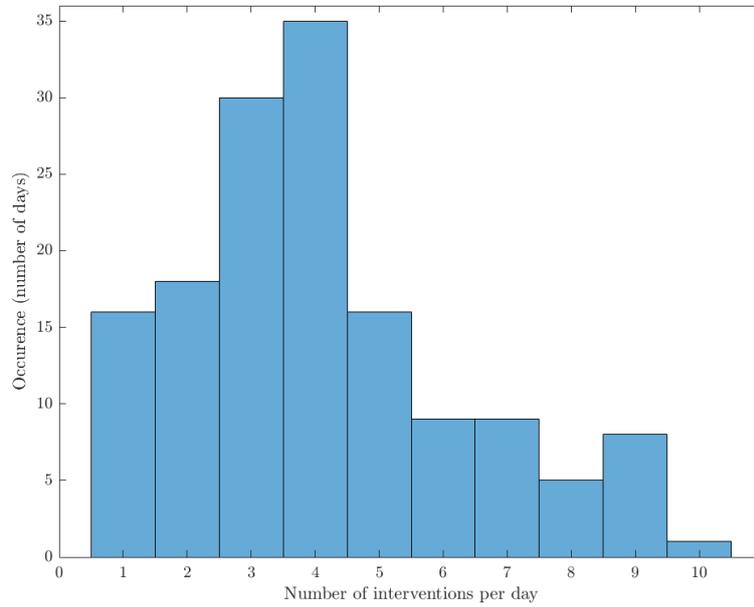


Figure B.1: A distribution of the number of CAG procedures per day.

C. Results of the RFE

Table C.1 shows the complete results of the RFE, based on the method presented in Section 3.3 and Equation (3.1). Based on observations from the RFE evaluation, the top 19 features were included in the final feature set.

Table C.1: The ranking of features based on the recursive feature elimination.

No.	Feature name	Score
1	Cumulative time	1.000
2	Acquisition frequency	0.476
3	Cumulative cine time	0.345
4	Table longitudinal	0.227
5	Patient age range	0.217
6	Object thickness	0.156
7	Long. distance travelled distance	0.139
8	Position detector	0.098
9	Table lateral	0.083
10	Frontal beam long. position	0.078
11	Rotation start	0.076
12	Rotation end	0.074
13	SID	0.073
14	Cumulative fluoro time	0.073
15	Frontal z-rotation	0.065
16	Position propeller	0.064
17	Run (kV)	0.064
18	Cumulative acquisition time	0.063
19	Number of lat. moves table	0.059
20	Cumulative non-acquisition time	0.059
21	Table height	0.058
22	Lat. distance travelled table	0.048
23	Angulation end	0.042
24	Number of cine uses	0.042
25	Angulation start	0.041
26	Position C-arm	0.039
27	Cumulative DAP	0.038
28	Run (mA)	0.035
29	Number of long. moves table	0.034
30	LCA phase	0.034
31	Cumulative AirKerma	0.032
32	Run (ms)	0.030
33	Vert. distance travelled table	0.030
34	AirKerma	0.028
35	DAP	0.028
36	Acquisition time	0.028
37	Number of fluoro uses	0.027
38	Procedure name	0.027
39	Requested dose	0.025
40	Acquisition number	0.025
41	Long. distance travelled front beam	0.025
42	RCA phase	0.025

No.	Feature name	Score
43	Time since acquisition	0.023
44	Number of long. moves front beam	0.023
45	Dose ratio	0.022
46	Number of vert. moves table	0.022
47	Frontal rotation detector	0.021
48	Application name	0.021
49	Cumulative DSA time	0.020
50	Number of DSA uses	0.020
51	Type	0.020
52	Frame speed	0.019

D. Complete Set of Logged Features

This section presents all parameters logged by the interventional X-ray system, including a brief description. All features that were left blank in the dataset were not investigated further and are therefore not defined in the tables below.

Table D.1: The original features logged by the interventional X-ray system.

No.	Feature name	Unit	Feature explanation
1	Exam	–	The procedure number within the dataset.
2	Date	–	The date on which the procedure took place.
3	Start time	<i>hh:mm:ss</i>	The time at which the acquisition was initiated.
4	End time	<i>hh:mm:ss</i>	The time at which the acquisition was completed.
5	Type	–	Acquisition type (fluoro, cine, DSA or Single-shot) [5, 24, 32, 48].
6	Channel	–	'Frontal' for all entries.
7	Road map	–	States whether the radiological Road Map technique was applied, using image subtraction at peak opacification [32].
8	Technique	–	Selected radiation technique to determine the hardware and software components used [32].
9	Application name	–	Type of application selected [37]. See Section 3.1.2.
10	Procedure name	–	Type of procedure selected [37]. See Section 3.1.2.
11	Fluo flavor	–	The selected level of fluoroscopy dose (low, medium, normal) [40].
12	Run	<i>kV</i>	Voltage across the X-ray tube. One of the X-ray generator settings [37].
13	Run	<i>mA</i>	Emission current from the X-ray tube. The higher this number, the higher the number of photons reaching the detector, thus the higher the signal intensity.
14	Run	<i>ms</i>	Pulse width of the X-rays.
15	Focus	–	The size of the focal spot used (small, large) [32].
16	Filter type	–	Type of filter used to reduce skin dose.
17	Cluster mode	–	Setting related to the hardware of the system.
18	Detector size	<i>mm²</i>	The selected area of the electron-receiving detector.
19	Detector diameter	<i>mm</i>	The selected diameter of the electron-receiving detector.
20	Acquisition time	<i>s</i>	Duration of the acquisition, referring to the total time during which X-rays were emitted by the tube.
21	Frame speed	<i>fps</i>	Capturing speed of the image detector [32].
22	Number of frames	–	The number of frames taken within the acquisition duration.
23	Sub images	–	The number of sub-images taken that encompass a catheter or arterial segment [47].
24	Phase 1 image speed	<i>fps</i>	Desired image speed for the first part of the procedure, manually selected by the cardiologist. Empty for most entries.
25	Phase 1 duration	<i>s</i>	Desired acquisition duration for the first part of the procedure, manually selected by the cardiologist. Empty for most entries.
26	Phase 2 image speed	<i>fps</i>	Desired image speed for the second part of the procedure, manually selected by the cardiologist. Empty for most entries.
27	Phase 2 duration	<i>s</i>	Desired acquisition duration for the second part of the procedure, manually selected by cardiologist. Empty for most entries.

Code	Feature name	Unit	Feature explanation
28	Phase 3 image speed	<i>fps</i>	Desired image speed for the third part of the procedure, manually selected by the cardiologist. Empty for most entries.
29	Phase 3 duration	<i>s</i>	Desired acquisition duration for the third part of the procedure, manually selected by cardiologist. Empty for most entries.
30	AGL TestShot DoseRatio	%	The dose ratio for the test phase. Empty for most entries.
31	TestShot	<i>kV</i>	The voltage across the X-ray tube during the test phase. Empty for most entries.
32	TestShot	<i>mA</i>	The emission current from the X-ray tube during the test phase. Empty for most entries.
33	TestShot	<i>ms</i>	The pulse width of the X-rays during the test phase. Empty for most entries.
34	Object Thickness	<i>cm</i>	Depth of the irradiated tissue. A larger bodypart requires a larger entrance dose to obtain the same detector dose [43]
35	Requested dose	<i>nGy</i>	Dose required by the detector to ensure sufficient image quality [43].
36	Dose ratio	%	Amount of X-ray dose measured by the image detector, with respect to the requested dose. [32, 43].
37	Entrance dose limit	<i>nGy</i>	The maximum AirKerma received.
38	Dose sense taste	–	Selected transfer curve (linear or logarithmic) applied to the image signal, manually selected by the cardiologist [32].
39	DAP	<i>mGy · cm²</i>	Dose Area Product. The dose of radiation delivered to the patient multiplied by the area of the exposed skin. Provides an estimated likelihood of skin damage.
40	Cumulative DAP	<i>mGy · cm²</i>	The sum of DAP values from all acquisitions, measured from the start of the procedure.
41	AirKerma	<i>mGy</i>	Measure of the amount of radiation energy deposited to a unit mass of air. Specifies the radiation concentration delivered to a point.
42	Cumulative AirKerma	<i>mGy</i>	The sum of AirKerma values from all acquisitions, measured from the start of the procedure.
43	MatrixSize	<i>px</i>	The desired size of the acquired image, in order to convert it using an analogue-to-digital device [12].
44	Shutter position	<i>μm</i>	Selects the area to be irradiated by changing the direction of the collimators [43].
45	Valid rectangle	<i>px</i>	The area to be irradiated, determined by the shutters and wedges.
46	Measuring field	<i>px</i>	Specifies the region of interested and affects image sharpness and radiation dose [32].
47	Wedge 1 distance	<i>μm</i>	The position of the first wedge. Wedges control the X-ray intensity by means of positioned filters [43].
48	Wedge 1 angle	°	The angle of the first wedge.
49	Wedge 2 distance	<i>μm</i>	The position of the second wedge.
50	Wedge 2 angle	°	The angle of the second wedge.
51	Angulation start	°	The roll of the C-arm at the start of the acquisition. See Figure 3.2.
52	Angulation end	°	The roll of the C-arm at the end of the acquisition.
53	Rotation start	°	The propeller of the C-arm at the start of the acquisition. See Figure 3.2.
54	Rotation end	°	The propeller of the C-arm at the end of the acquisition.
55	SID	<i>m</i>	Source-to-image distance. The distance between the X-ray tube and the detector.

Code	Feature name	Unit	Feature explanation
56	Position C-arm	$^{\circ} \times 10^2$	The unfiltered angulation of the C-arm.
57	Position detector	μm	The vertical position of the detector. Unfiltered version of the SID. See Figure 3.2.
58	Position propeller	$^{\circ} \times 10^2$	The unfiltered rotation of the C-arm.
59	Frontal beam longitudinal	μm	The longitudinal position of the X-ray system. See Figure 3.2.
60	Frontal beam transversal	μm	Zero for all entries.
61	Frontal rotate detector	$^{\circ}$	The position of the detector, moving between portrait and landscape position. See Figure 3.2.
62	Frontal swing	$^{\circ}$	'Unknown' for all entries.
63	Frontal Z-rotation		The swing of the C-arm. See Figure 3.2.
64	Table height	μm	The vertical position of the operating table. See Figure 3.2.
65	Table lateral	μm	The lateral position of the operating table. See Figure 3.2.
66	Table longitudinal	μm	The longitudinal position of the operating table. See Figure 3.2.
67	Contrast TransferTaste	–	Empty for all entries.
68	Subtraction contrast TransferTaste	–	Empty for all entries.
69	Perception TransferTaste	–	Empty for all entries.
70	Processed export TransferTaste	–	Empty for all entries.
71	SmartMask contrast TransferTaste	–	Empty for all entries.
72	Motion comp. active	–	Empty for all entries.
73	Motion comp. taste	–	Empty for all entries.
74	Spatial NR active	–	Empty for all entries.
75	Spatial NR taste	–	Empty for all entries.
76	Temporal NR active	–	Empty for all entries.
77	Temporal NR taste	–	Empty for all entries.
78	Temporal NR factor	–	Empty for all entries.
79	Temporal NR factor GeoMoving	–	Empty for all entries.
80	Temporal GeoMoving applicable	–	Empty for all entries.
81	Temporal NR factor OutOfSteam	–	Empty for all entries.
82	Subtraction active	–	Empty for all entries.
83	Start subtraction index	–	Empty for all entries.
84	Build subtraction mask	–	Empty for all entries.
85	Subtraction MaskTaste	–	Empty for all entries.
86	StartBuild subtraction MaskIndex	–	Empty for all entries.
87	Build subtraction MaskDuration	–	Empty for all entries.
88	Build landmark mask	–	Empty for all entries.
89	Build trace mask	–	Empty for all entries.
90	Trace type	–	Empty for all entries.
91	Contrast	–	Empty for all entries.
92	Brightness	–	Empty for all entries.

Code	Feature name	Unit	Feature explanation
93	Edge enhance	–	Empty for all entries.
94	Edge enhance kernel	–	Empty for all entries.
95	Harmoni zation	–	Empty for all entries.
96	Video invert	–	Empty for all entries.
97	Subtraction contrast	–	Empty for all entries.
98	Subtraction brightness	–	Empty for all entries.
99	Subtraction Edge-Enhance	–	Empty for all entries.
100	Subtraction Edge-Enhance kernel	–	Empty for all entries.
101	Subtraction harmoni zation	–	Empty for all entries.
102	Subtraction video invert	–	Empty for all entries.
103	Zoom	–	Empty for all entries.
104	Pan X	–	Empty for all entries.
105	Pan Y	–	Empty for all entries.
106	Subtract enabled	–	Empty for all entries.
107	Subtraction gain	–	Empty for all entries.
108	Live gain	–	Empty for all entries.
109	Subtraction live gain	–	Empty for all entries.
110	Landmark enabled	–	Empty for all entries.
111	Landmark gain	–	Empty for all entries.
112	Vessel gain	–	Empty for all entries.
113	PixelShift X	–	Empty for all entries.
114	PixelShift Y	–	Empty for all entries.
115	NonSubtract device	–	Empty for all entries.
116	Hospital	–	The hospital at which the procedure took place.
117	System	–	Empty for all entries.
118	Exam ID	–	Anonymous procedure identification number.
119	Patient ID	–	Anonymous patient identification number.
120	Physician ID	–	Anonymous physician identification number.
121	Patient age from	<i>years</i>	Minimum year of the patient’s age range, divided up into 10-year segments.
122	Patient age till	<i>years</i>	Maximum year of the patient’s age range, divided up into 10-year segments.
123	Patient weight from	<i>kg</i>	Minimum weight of patient’s weight range, divided up into 10-kilogram segments. Not logged correctly.
124	Patient weight till	<i>kg</i>	Maximum weight of patient’s weight range, divided up into 10-kilogram segments. Not logged correctly.
125	FXD	–	Empty for all entries.