Vilde Wøien

Supervised Deep Learning for Perioperative Cardiovascular Monitoring

June 2022

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology
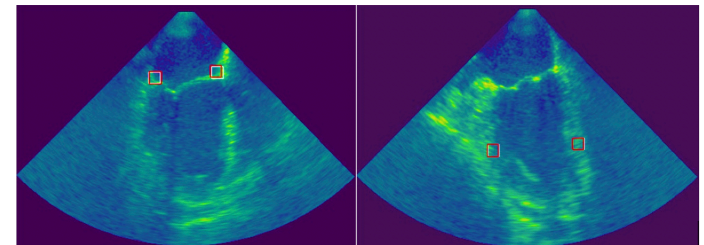
NTNU
Norwegian University of
Science and Technology

Supervised Deep Learning for Perioperative Cardiovascular Monitoring

Vilde Wøien

Engineering and ICT
Submission date: June 2022
Supervisor: Jørn Vatn
Co-supervisor: Gabriel Kiss and Anders Austlid Taskén

Norwegian University of Science and Technology
Department of Manufacturing and Civil Engineering

Preface

This paper is a Master's thesis at Norwegian University of Science and Technology (NTNU) as part of the Engineering and ICT study program with the main profile ICT and Operation Management, carried out in the spring of 2022. This study is based on the preliminary research written during the autumn semester in 2021 for the course named Production Management, Specialization Project -TPK4530. The supervisors work at NTNU and St. Olavs Hospital, Trondheim University Hospital. The project was chosen due to my interest in the intersection between medicine and technology and digital transformations.

Acknowledgements

Sammendrag

På grunn av høy risiko for kardiovaskulære komplikasjoner under hjertekirurgi, overvåker leger pasientens hjerte i den perioperative fasen. Ekkokardiografi er mye brukt til å vurdere hjertefunksjon og hemodynamikk. Overvåking av hjertefunksjon og hemodynamikk består vanligvis av målinger av blodtrykk, hjertefrekvens og oksygenmetning. Vurdering av ultralydbildene er en kvalitativ visuell inspeksjon utført av eksperter. Følgelig har evalueringen en tendens til å ha inter- og intravariabilitet. I tillegg er det for tidkrevende for en kardiolog eller anestesilege å evaluere ultralyd regelmessig. Derfor er automatiserte tiltak ønskelig. Transøsofagal ekkokardiografi har vist seg nyttig for hjertekirurgi, men det er mangel på programvare for automatisk estimering av kvantitative hjertefunksjonsmål, for eksempel myokardiell strain (deformasjon av hjerteveggen).

Denne oppgaven har som mål å spore punkter i hjertet for å automatisk estimere myokardiell strain, som beskriver regional hjertefunksjon. En veiledet (eng: supervised) dyp læringsmetode, kalt TransformerTrack, ble tilpasset videosekvenser, tatt med transøsofagal ekkokardiografi, for å spore punkter i hjertestrukturen. TransformerTrack er designet for å utnytte den tidsmessige informasjonen blant videobildene for nøyaktig sporing. Metoden er enda ikke brukt til medisinske oppgaver.

TransformerTrack produserte lovende sporingsresultater på sporing av punkter modellen ble trent på. Modellen predikerte posisjonen til punktene i hjertet med en gjennomsnittlig avstand på 5.565 +− 4.763 piksler mellom prediksjon- og referansepunktet, på et bilde med størrelse på ca. 350x500 piksler. Modellen viste en dårligere evne til å spore punkter den ikke ble trent på, og videre arbeid ble foreslått for å forbedre dette.

Arbeidet i denne oppgaven er en del av et større forskningsprosjekt som utvikler et automatisert overvåkingsverktøy av hjertefunksjoner. Det ble gjennomført en As-Is-analyse av dagens arbeidsprosesser i den perioperative fasen basert på intervjuer med leger og et litteratursøk. I tillegg ble fordeler og ulemper ved å bruke maskinlæring til medisinske oppgaver undersøkt. En To-Be prosesskartlegging ble utformet for å bidra til implementasjon av det automatiserte overvåkingsverktøyet i operasjonsrommet. I tillegg bidro kartleggingen til å vurdere relevansen av de vanlige problemene knyttet til medisinsk maskinlæring. Kartleggingene viste at et automatisert overvåkingsverktøy vil påvirke kommunikasjonen mellom leger, arbeidsflyten og pasientsikkerheten. I utviklingen av verktøyet blir det en avveining mellom falske positive og falske negative utfall basert på pasientsikkerhet og sykehusets ressurser.

Abstract

Due to a high risk of cardiovascular complications for patients undergoing cardiac surgery, the physicians monitor the heart in the perioperative phase. Echocardiography is widely used to assess cardiac function and hemodynamics. Monitoring of cardiac function and hemodynamics usually consists of measurements of blood pressure, heart rate, and arterial oxygen saturation. The assessment of the ultrasound images is a qualitative visual inspection performed by experts. Consequently, the evaluation tends to have inter-and intravariability. Additionally, it is too time-consuming for a cardiologist or anaesthesiologist to evaluate ultrasound regularly. Hence, automated measures are beneficial. Transesophageal echocardiography has proven helpful for cardiac surgery purposes, but there is a lack of software for automatic estimation of quantitative cardiac function measures, such as longitudinal strain.

This paper aims to track points in the heart to contribute to automatic longitudinal strain estimation, describing regional cardiac function. A supervised deep learning method called TransformerTrack was adapted to transesophageal echocardiography video frames to track points in the cardiac structure. TransformerTrack is designed to utilize the temporal information among video frames for accurate tracking and is currently not adapted to the medical environment.

A quantitative assessment showed that the deep learning method produced promising tracking results with a mean distance of 5.565 +− 4.73 pixels between the predicted- and reference landmark, on a frame of approximately 350x500 pixels. The model showed a poor ability to generalize, and further work was suggested for mitigation.

The work in this thesis is part of a larger research project that develops an automated monitoring tool for cardiac functions. An As-Is analysis of the current work processes in the perioperative phase was conducted based on interviews with doctors and a literature review. Additionally, common advantages and disadvantages of medical machine learning were investigated. A To-Be process mapping was proposed to contribute to implementing the automated monitoring tool in existing work processes and examining the common problems' relevance. The mappings revealed that the tool would affect the communication between physicians, the workflow, and patient safety. In the development of the tool, there will be a trade-off between false positive and false negative outputs based on patient safety and the hospital's resources.

Contents

Figures

Tables

Acronyms

SCM supply chain management. 15

STE speckle tracking echocardiography. 2

TDI tissue doppler imaging. 2

TEE transesophageal echocardiogram. xii, 2–4, 12, 13, 29, 31, 36, 39, 43–46, 53–55, 59

TN true negative. 44, 58

TP true positive. 44, 58

TTE transthoracic echocardiogram. 13, 54

VSM value stream mapping. 14, 15

Chapter 1

Introduction

1.1 Background

Cardiovascular diseases are widespread globally [1]. Between 1% and 2% of the adult population in the western world have heart failure, and 10% of those over 70 years are affected by it [2]. Cardiac surgery entails a risk of complications, depending on both the nature of the procedure and the patient's condition before the operation [3]. Interventions such as vascular surgery, bypass surgery, and valve-related interventions often affect cardiac performance [4]. Cardiac dysfunction such as decreased myocardial contractility, atrial fibrillation, and myocardial infarction are common consequences of cardiac surgery [5]. Complications can result in mortality, increased length of stay, and increased costs [6]. Even patients undergoing noncardiac surgery are prone to cardiovascular complications during the perioperative period, and these complications account for the majority of the cause of post-operative morbidity and mortality [7].

The patient's cardiac function is monitored in the perioperative phase to reduce the risk of complications [6, 8]. The patient's vital signs are continuously monitored, including blood pressure, heart rate, and blood oxygen saturation. Additionally, a visual inspection of ultrasonic imaging of the heart is completed [9]. Echocardiography is the most common imaging modality in clinical application [10]. It is an indispensable tool for assessing cardiac function because it allows for the visualization of the important structures of the human heart. Digital ultrasound imaging gives doctors an insight into cardiac anatomy and function [11]. The echocardiographic assessment of cardiac function is a task reserved for specialists [12]. When assessing the ultrasound images, the specialists inspect if the heart muscle and the valves are working as they should [13]. Traditionally, visual estimation of myocardial morphophysiology has been the standard of cardiac contraction evaluation [14].

Manual assessment of echocardiography is qualitative, resulting in a high tendency of inter- and intravariability [14, 15]. Therefore, automation of perioperative echocardiographic monitoring is desired [4]. There exists several quantitative indicators measuring cardiac function. *Myocardial strain* is such an indicator, describing regional cardiac function. Myocardial strain is a measure of the deformation of a segment in the cardiac wall tissue [16]. Myocardial strain

imaging is applied in several clinical settings [16]. Imaging modalities for myocardial strain estimation has been tested and compared, and various methods for myocardial strain analysis have been tested [16–19].

There is lacking a standard method for assessment of strain in vivo [20]. Currently, the most used methods for myocardial strain estimation are types of myocardial velocity imaging, for instance tissue doppler imaging (TDI) [21–24]. TDI provide information on the myocardial deformations by measuring the velocity of the myocardium [25]. The velocity measurement is completed when the beam of the ultrasound probe is parallel to the tissue to be measured. Therefore the method is limited by its angle dependency [25, 26]. Research has been conducted to overcome this limitation, and speckle tracking echocardiography (STE) has proven to be an angle independent method feasible for measuring myocardial strain in real-time [26]. STE uses the natural occurring speckles in the ultrasound images to track the motion of the cardiac wall tissue [20]. STE is limited by poor temporal resolution [27]. The strain assessment is semi-automated because a manual definition of cardiac anatomy is required. Myocardial strain derived from TDI, and STE results in earlier detection of cardiac dysfunction than the traditional methods. The fully automatic strain estimation is undergoing development [15].

Deep learning (DL) is a subfield of machine learning (ML), and is a prevailing tool in the field of computer vision (CV). CV tasks, such as object detection and action recognition, focuses on extracting information from digital images [28]. CV has proven to successfully automate tasks related to image analysis in a wide range of environments, including medical image analysis [29]. Measures of left ventricle (LV) functions, like myocardial strain, is currently done based on digital imaging [22, 26, 30]. Therefore, CV could be feasible for the automatic strain estimation purpose [27]. Visual object tracking is a CV task aiming to automatically obtain the location of an object in each frame of a video sequence. Visual object tracking is an active research area with a wide range of applications, for instance, surveillance and human-computer interactions [31–34].

With the rapid growth of ML as a technological tool, ML methods are used for clinical decision support in radiology, dermatology, cardiology, and pathology [35]. For instance, Bouton et al. [36] developed an ML system to restore the control of movement in patients with a particular form of paralysis. The rise of CV as a research field has contributed to increasing the range of tasks ML can support due to the amount of patient information held in clinical images. Esteva et al. [37] have successfully used CV to identify skin cancer subtypes. Visual object tracking has successfully been applied on ultrasound images. DL models have been tested on the identification of nerves in ultrasound images for the support of clinical decision-making [38].

Transesophageal echocardiography (TEE) is a widespread type of echocardiogram [39, 40]. TEE has proven to be useful for several cardiac surgery purposes [4]. Minimal invasive cardiac surgeries, such as heart valve surgeries, are often performed under TEE monitoring. The adaption of TEE monitoring purposes in the operating room (OR) is limited due to the lack of software tools and requirements for extensive user interaction by an echocardiographer expert

[4]. A specialist must be called pre- and post surgery to clinically assess the echocardiographies [4, 8, 12, 41].

There is an ongoing project involving the Norwegian University of Science and Technology and St. Olavs Hospital, Trondheim University Hospital, called "Automated machine learning based perioperative monitoring of cardiac function (APOC)". The project creates a solution that solves the problems related to the manual assessment of echocardiography and the lack of software to utilize TEE images. The principal goal of the research group is to develop a functional prototype of an automated ML based tool for cardiac function analysis based on TEE images *(further called "tool" or "automated tool")*. The tool aims to solve the challenges mentioned above by continuously measuring several heart function metrics in real-time during the perioperative phase based on TEE images [4]. One of the cardiac measures that the monitoring tool will base its output on is the longitudinal strain [42, 43]. The work in this thesis on visual object tracking of TEE images is a part of the APOC project.

A well-performing clinical ML model is not necessarily adequate for the tool to be routinely utilized and endorsed by physicians [44]. Recent studies have shown that the implementation of ML methods in health care fails due to poor integration with the clinical workflow and the staff [45, 46]. For successful implementation and sustainable use of new tools, both research regarding the technical performance, potential practical challenges, and comparison of current and future work processes is essential [47]. Developers' knowledge of how the work processes are today and how the tool is going to change them is crucial for better software [48].

1.2 Objectives

The main objectives of this master's thesis are:

- Map current work processes in the OR and propose a future state mapping with the implemented monitoring tool
- Investigate if the benefits of the mappings will contribute to the development and integration of the tool
- Investigate if the current challenges and advantages of current use of AI in health care applies to the monitoring tool
- Adapt TransformerTrack and TEE images for visual object tracking

1.3 Research Questions

The research questions (RQs) are as follows:

- **RQ1**: How will the monitoring tool affect the processes in the perioperative phase?
- **RQ2**: How will the adapted visual tracking model, TransformerTrack, perform on TEE images?

1.4   Approach

Interviews with doctors at St. Olavs Hospital are carried out to map the current work processes and propose a future state. An As-Is analysis is executed, and a To-Be process mapping is proposed to define the future state of the work processes in the OR when the functional tool is developed by the APOC project is in place. A literature review is conducted to investigate the current benefits and difficulties regarding the implementation of ML-based methods in health care, specifically for cardiac purposes. The questions in the interviews are based on the findings in the literature review. Additionally, the usefulness of the process mappings is discussed in chapter 5 in light of the findings of the literature review. Figure 1.1 illustrates the approach.
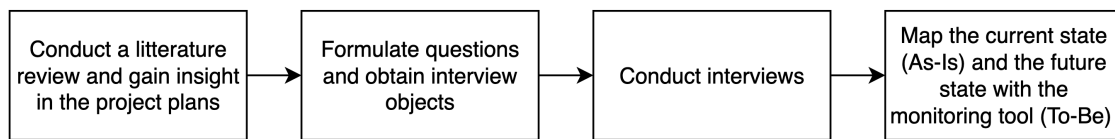
Figure 1.1: A simple illustration of the approach to analyze the current and future work processes.

CV is used to contribute to the APOC project's by developing an automated cardiac monitoring tool. More specifically, a visual object tracking method is applied on ultrasound images with the aim to estimate longitudinal myocardial strain *(from now on called strain or longitudinal strain)*. The tracking of points in the heart provides the position of the points at a given time, which makes it possible to automatically calculate strain based on TEE images. Therefore, a DL model is trained to track certain points in the cardiac structure and measure its performance. Figure 1.2 illustrates the approach to the automatic strain estimation using DL.

The TransformerTrack method is used for visual tracking of TEE images to contribute to the automation of strain estimation. The technique differs from most tracking by focusing on the temporal relationship among successive frames in the video sequence [49]. Rich temporal contexts exist across successive frames of a cardiac cycle ultrasound video sequence. Additionally, TransformerTrack has proven to produce consistent tracking of a specified object in video frames containing distracting objects [49]. This model characteristic is beneficial to strain estimation since similar areas in the heart structure can confuse a tracking model. TransformerTrack has not yet been used for medical imaging purposes and requires development and adjustment to fit the presented problem. By utilizing the TEE images as input to the model, there is a good chance that the output will be reliable and reproducible [50].

The approach aims to measure the tracking performance of the model on the TEE images. DL tracking methods are frequently used for digital imaging purposes, and there exist several variants. Thus, it is essential for the APOC project to adapt a state-of-the-art model to the problem and analyze the performance [51].
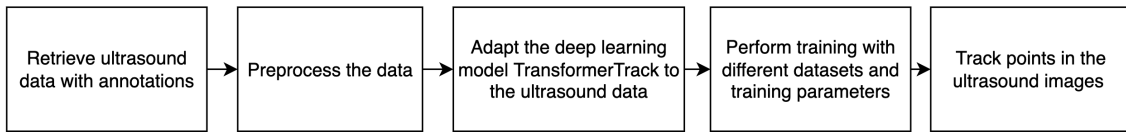
Figure 1.2: A simple illustration of the approach to contribute to the automatic strain estimation.

1.5 Limitations, Scope and Target Group

Since the APOC project is in a prototype phase, the final details about the implementation of the tool are not formally decided. Further investigation is needed to gain insight into the final goals of the monitoring tools and secure the thoroughness of the analysis.

CV in clinical care can provide multiple advantages, including increased precision and efficiency in diagnostics and treatment, improved planning, enriched opportunities for research, and enhanced sustainability of services. The use of AI creates new challenges regarding security, privacy, regulation, accountability, ethics, and management [52]. The work process mapping analysis' scope is the processes in the OR, the most relevant stakeholders in the OR, and the time scope is the perioperative phase. The scope of visual object tracking is measuring training and testing performance. Key focus areas are performance, safety, communication, and process time.

The target group of the thesis is those with an Engineering and ICT master's degree. Those who have a specialization in AI and ML can skip the theory chapter concerning these topics in section 2.6. A short definition of relevant AI concepts for visual object tracking of ultrasound images is presented in table C.1 in appendix C. Nearly the whole theory chapter 2 and subsection 3.2.1 are from the specialization project due to their relevance.

1.6 Outline

The remaining of the report is organized as follows:

- **Chapter 1. Introduction:** A presentation of the motivation of using ML for automatic monitoring of cardiac function. The current approaches to heart monitoring during cardiac surgery is presented. Additionally, this chapter contains suggestions to solve the current problems based on state-of-the-art ML models and previous attempts to use ultrasound imaging for cardiac surgery monitoring.
- **Chapter 2. Theoretical Background and Review of the Literature:** A presentation of the findings of a literature review and the theoretical background required for profound insight into the topics. These topics are ML, the human heart, ultrasound, production management (PM), digitization, ML in health care, the effect of ML on work processes, and value trade-offs in the health sector.
- **Chapter 3. Data and Method:** A presentation of the medical data used in the study, its metadata, the data retrieval, data annotation, and preprocessing. Additionally, an

explanation of the methods, literature review, process analysis, and the DL model.

- **Chapter 4. Results:** The results of the interviews, process mappings, and the visual object tracking.
- **Chapter 5. Discussion and Further Work:** The findings of the study and an analysis of the results are presented and discussed. The main focus is to answer the research questions. Further work and limitations of the work are also presented.
- **Chapter 6. Conclusion:** The conclusion of the study is drawn.
- **Bibliography:** Overview of all the referenced papers and books.
- **Appendix A. Technical Commands:** The commands for running the model.
- **Appendix B. Libraries and Packages:** The libraries and packages used for visual object tracking.
- **Appendix C. Concepts of Artificial Intelligence:** An overview of some relevant concepts of AI related to visual object tracking of ultrasound images. Useful for readers new to AI to look up short definitions.
- **Appendix D. Interview:** An overview of the questions the interviews are based on.

Chapter 2

Theoretical Background and Review of the Literature

Firstly, the anatomy and the functions of the heart is presented with the existing ultrasound technology. These two topics cover the essential theory, and the thesis do not contribute to these areas. However, the heart and ultrasound background theory is necessary for understanding and developing the automatic monitoring tool.

The literature review focus on production management (PM) and digitization. PM has developed techniques for analyzing and understanding work processes to improve these processes, such as As-Is and To-Be. The literature review investigate what has been done in other comparable areas regarding how digitization works or its challenges. The digital tool at the "process level" is a tool that can make the difference between life and death. Therefore, it is essential to investigate the implementation and the effects of decision-making in health care.

The ML fundamentals is presented. The theory of ML fundamentals is presented to substantiate the method behind the monitoring tool used for strain estimation. The general ML theory is initially presented, and the state-of-the-art methods are described along with the relevant theory. Lastly, the advantages and challenges of implementing ML in the cardiac domain are investigated.

Figure 2.1 shows the used search engines and key search words related to each topic of the literature review. Search words were chosen based on the research questions, theory, and frequently used words in the relevant literature.

Topic	Search engines	Key search words
Machine learning	Scopus, Google Scholar, Oria	AI-, ANN-, CNN, CV- and ML fundamentals, training/testing ML models, state-of-the-art models, visual tracking, transformers, attention mechanism, backpropagation, convolutions, pooling, linear rectifiers, hyperparameters, biological neural networks, gradient descent, implementation, data pre-processing,
Digital tools and work processes	Scopus, Google Scholar, Oria	Smart PPC, industry 4.0, digital technologies, food manufacturing, safety, quality, resource efficiency, cost reduction, customer satisfaction, digitization, digitalization, digital transformation, effectiveness, work processes, improvements,
Heart	PubMed, Scopus, Google Scholar, Oria	Cardiac anatomy, cardiac function, human heart, cardiovascular deceases, left ventricle, monitoring, ultrasound, TEE, TTE, patient care, risks, MAPSE
AI in health care	PubMed, Scopus, Google Scholar, Oria	AI in health care, decision support, automation, digitalization in health care, digital assessment, digital monitoring,

Figure 2.1: Overview of the search engines and key search words used in the literature review

2.1   The Anatomy of the Human Heart

The human heart is a hollow muscular organ responsible for the bloodstream in the body [53]. It is common to distinguish the right and the left section of the heart, which each consists of an upper chamber, the *atrium*, and a lower chamber, the *ventricle*, as illustrated in figure 2.2 [54]. The function of the ventricles is to work like pumps, and the atrium chambers receive blood that returns to the heart. The atrium and the ventricle are separated, on both sides, by connective tissue, *annulus fibrosus*. Annulus fibrosus consists of four connected rings, which form four valves constructed for the blood to flow in one direction. The *mitral valve*, separates

the left ventricle (LV) and the left atrium, the *tricuspid valve* separates the right ventricle (RV) from the right atrium. The *aortic valve* separates the LV and the main artery, *the aorta*, and the *pulmonary valve* separates the RV and the *pulmonary artery* [53].

The wall of the heart is composed of three layers named *epicardium*, *myocardium* and *endocardium*. The endocardium is a thin inner surface lining the heart's interior, consisting of endothelium and connective tissue. The myocardium constitutes the middle layer and most of the heart wall and is a cardiac muscle with cells responsible for the contraction of the heart, *cardiomyocytes*. The epicardium is the outer layer of the heart wall, composed of a thin serous layer and a subserous layer. The subserous layer is embedded with coronary blood vessels for the supply of oxygen-rich blood and nerves [54].
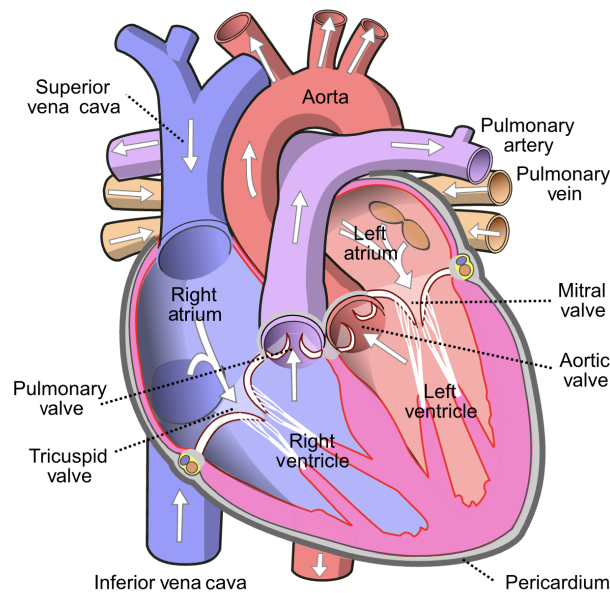
Figure 2.2: The heart is illustrated with names and blood flow. White arrows indicated the direction of the blood flow. Illustration by Wikimedia user Wapcaplet[1], reproduced under the CC BY-SA 3.0 license [2]

The right section of the heart pumps the blood through the pulmonary circuit, and the left section pumps it through the systemic circuit. The right atrium receives venous blood from the inferior and superior vena cava. The blood flows from the right atrium to the RV, which pumps the blood through the pulmonary artery to the lungs. The deoxygenated blood is added oxygen in lungs, and the carbon dioxide in the blood is released into the lungs. The blood returns to the left atrium of the heart via pulmonary veins from the lungs. The blood streams from the left atrium to the left ventricle. Then the blood is distributed through the aorta to all the organs of the body [53].

The heart contracts periodically to create pressure differences in the circulation system to support the bloodstream. The contraction phase of the heart, the *ventricular systole,* and the relaxation phase, the *ventricular diastole*, compose the cardiac cycle, illustrated in figure 2.3.
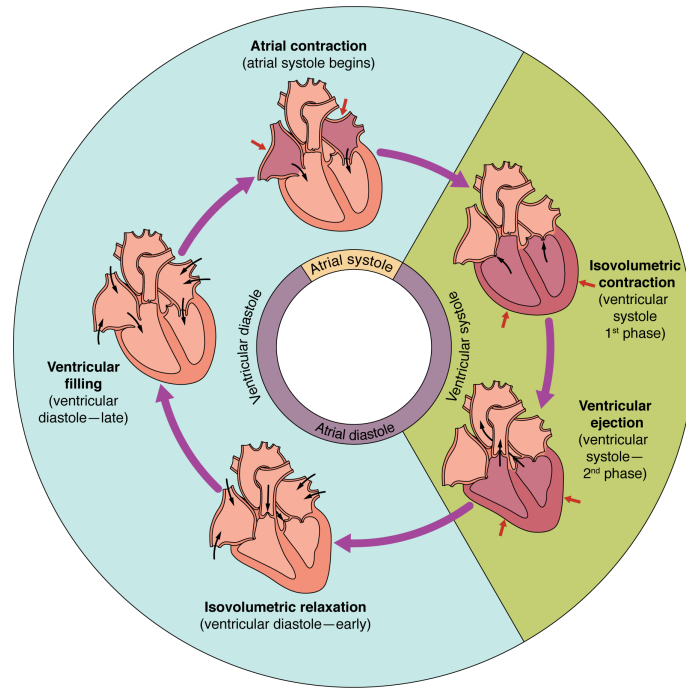
Figure 2.3: The cardiac cycle [55].

During invasive surgeries it is common to use a cardiopulmonary bypass (CPB) machine to temporarily stop the cardiac contraction. The machine lets the surgeon stop the patient's heart while maintaining blood circulation [56]. The doctors drain venous blood from the body to a reservoir, then it is oxygenated and pumped back into the patient's arterial system [56].

2.2   Cardiac Function

Identification of LV dysfunction is important when assessing the cardiac function, since heart failure often occurs as a consequence of left ventricular dysfunction [51]. Quantification of the cardiac function is useful for physicians to identify abnormalities [17]. LV myocardial strain quantitatively describes the function of the LV by the wall motions [16, 51]. Myocardial strain is defined by measuring a local or global shortening, thickening, or elongation of the myocardium in one or more dimensions as the ventricle contracts [25]. Figure 2.4 shows a graphic representation of the principal deformations of the myocardium; longitudinal strain, radial strain, circumferential strain [17].

Strain can be estimated based on echocardiographic images [17]. The deformation can be expressed relative to the initial length, which is called *Lagrangian strain* [22]. Given two points
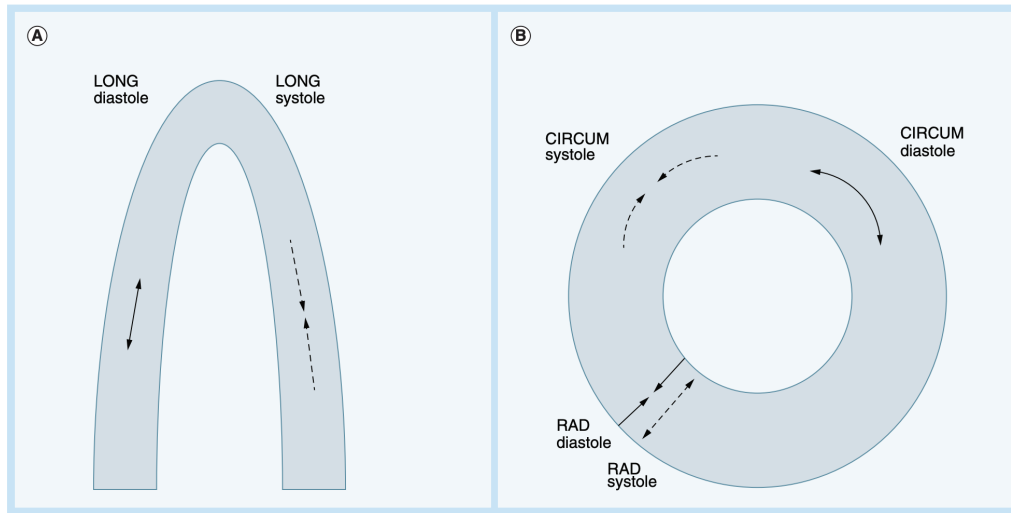
Figure 2.4: Illustration of the principal deformations of the myocardium. **(A)** Long axis view of LV. **(B)** Short axis view of LV. LONG: Longitudinal strain, RAD: Radial strain, CIRCUM: Circumferential strain. © 2012 by American Heart Association

in a specified region of the myocardium, the myocardial Lagrangian strain, $\epsilon(t)$, can be written as,

$$\epsilon(t) = \frac{L(t) - L_0}{L_0} \tag{2.1}$$

where t is the time, and $L(t)$ is the distance between the two points at a given time, illustrated in 2.5. L_0 is the initial distance, thus $L_0 = L(t_0)$ [15]. By convention, the strain is defined such that lengthening is represented as a positive strain value and shortening has a negative value, as in 2.1 [22]. The quantification of myocardial strain using echochardiograpy provides information which is useful in a variety of clinical settings [20, 57–59].
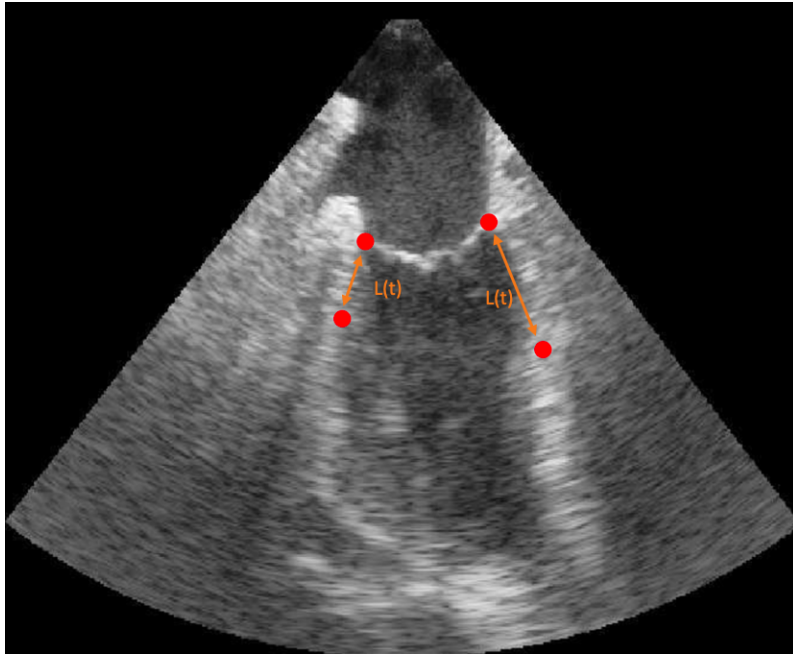
Figure 2.5: Illustration of the distance between the points used for strain computation. The image is a 2D TEE image. The red dots are the landmark annotations of the mitral points (the upper points) and points at the myocardium (the lower points). The strain equation is presented in equation 2.1.

2.3 Ultrasound

Ultrasound is a principal approach used in digital image processing and can operate on the electromagnetic spectrum outside of the human perception. Ultrasound imaging has a varied field of application and is best known for its use in obstetrics [60]. The technique is essential for clinical diagnoses and patient monitoring [61]. The ultrasound technique can produce different kinds of images, and this paper will focus on B-mode images [60]. B-mode ultrasound is one of the most widespread due to its low health risk, and real-time availability [62].

The ultrasound system consists of a computer, an ultrasound probe transmitting and receiving sound waves, and a display. The ultrasound images are generated by this step-by-step approach [60]:

1. The ultrasound transducer transmits high-frequency, 1-5 MHz, sound pulses into the body.
2. Sound pulses travel as sound waves until it hits tissue or organ boundaries. A boundary can, for instance, be between soft tissue and bone. When the sound waves hit a boundary, a part of it is reflected in the transducer as an echo.
3. The transducer receives reflected sound waves, and then the sound waves are converted to digital signals and relayed to a computer.
4. The computer uses the assumption of speed of sound, c, in area of interest, and time, t,

to calculate the distance d from the probe to the boundaries as $d = ct/2$

5. The computed distances and intensities of the echoes form a two-dimensional image [60]. The brightness of the image is related to the amplitude, the height of the ultrasound wave, of the echo [61]. Thus, these images are named B-mode images, where B is for brightness [60].

For one ultrasound image, millions of sound pulses are transmitted each second [60].

Echocardiography is cardiac ultrasound imaging and assists decision-making in several clinical settings. Echocardiography performs well at depicting cardiac structures, and the B-mode images resemble the anatomy of the heart [61]. The ultrasound transducer can be moved, placed, and angled to obtain various views [60]. In transthoracic echocardiogram (TTE) the transducer is manually moved on the exterior of the chest and abdominal wall and is an available and versatile procedure. This examination is noninvasive and accessible. However, other organs or bones can disturb the echo and contribute to inaccuracies [63]. In TEE, the transducer is placed inside the esophagus, which is close to the heart. This method is, in contrast to TTE, invasive, and the placement of the transducer can lead to oropharyngeal-, esophageal- or gastric trauma [39]. TEE is in many cardiac surgical cases more practical to use than TTE [64]. The main advantage of TEE is that once the probe is in place, it is assumed to be more or less stable, and it does not require an echocardiographer to acquire images continuously. The placement of the transducer during the TEE and TTE approach is illustrated in figure 2.6.

The transducer can be positioned in different ways, resulting in different images. The most common is the four-chamber view, and the structures visible in this view are the right ventricle (RV), left ventricle (LV), right atrium (RA), and left atrium (LA). The two-chamber view images show the left ventricle (LV) and the left atrium (LA). The displays of the four-chamber view and two-chamber view is illustrated in figure 2.7 [65].

Standard or Transthoracic Echo   Transesophageal Echo or TEE
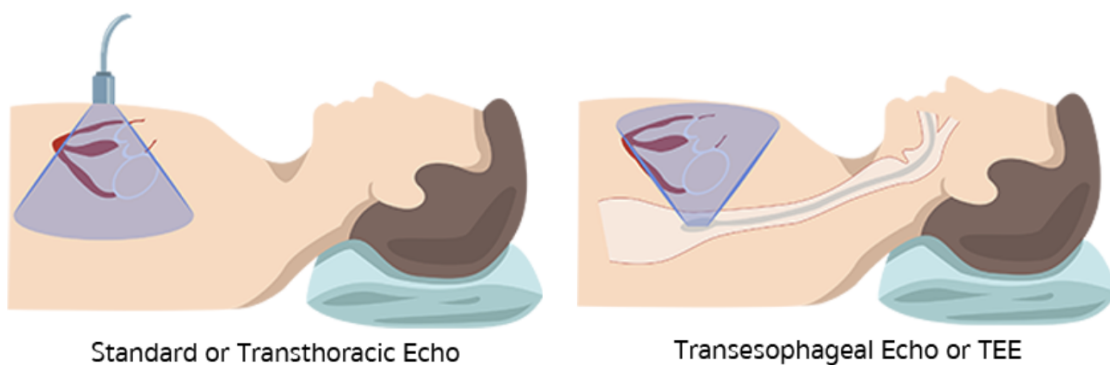
Figure 2.6: The left figure illustrates the transducer placement for TTE; on the outside of the chest or abdominal wall. The right figure illustrates the transducer inside the esophagus for TEE [66].
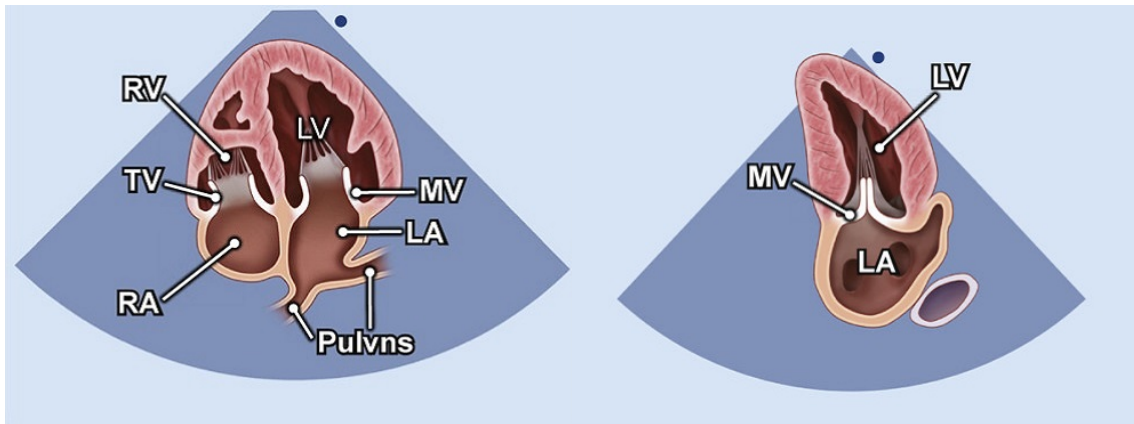
Figure 2.7: Four-chamber view and two-chamber view, respectively. Right ventricle (RV), left ventricle (LV), right atrium (RA) and left atrium (LA), tricuspid valve (TV), mitral valve (MV), pulmonary veins (Pulvns) [67].

2.4 Production Management and Digitalization

PM focuses on how organizations manage to deliver to their customers at the right time, cost, and quality [47]. Process productivity results from the interaction between people, technology, and processes. Some elements from PM are central when planning and conducting surgeries. The goal of PM is to manage the business processes that transform input into output, in the form of products and services [47]. The critical input factors regarding PM in the OR consist of health professionals, technology, machinery, patients, and equipment. The *customers*, in a medical PM aspect, are the patients. The management of the processes during a surgery aims to optimize the safety of the patients, and the efficiency regarding the use of time, equipment, and physicians. Management models, such as As-Is and To-Be models, are an abstraction of the value chain that shows how production and logistics are organized and managed [68, 69]. It is an advantage to use these models when planning the work in the OR, since strategic planning might lead to improved time use, fewer resources, and improved safety. The processes, material flow, information flow, and decision management should then be mapped and analyzed [70].

Digitization is the process of changing something from analog to digital. Digitalization is the use of digital technologies and digitized data to improve processes, functions, and activities, and the aim is to benefit from the change of the processes [71]. With the growth of digitalization, the need for analysis tools to optimize the implementation of digital tools has increased. The PM tools As-Is and To-Be process mappings are helpful tools for the implementation of the digital tools [68].

In PM, there are several techniques available to analyze work processes, and they are often used related to digitization processes. The techniques are applied to pursue the goal of understanding digitization performance. Value stream mapping (VSM) is one technique to map the process's current status, mainly used to map the flow of one product through a manufacturing

system [72]. VSM is applied to health care and other sectors to introduce the management philosophy of lean manufacturing to facilitate continuous improvements [73]. An aim with VSM in health care is to improve key performance operational indicators such as cost, clinical quality, patient safety, and process time. There are several factors affecting the performance of digitization, and therefore the *radar chart* is commonly used to visualize and evaluate operations [74]. The radar chart provides a simplified representation of the main KPIs measures, however, the timely aspect is not included in the chart [75]. The As-Is and To-Be process mappings can be in flow-chart format, where key participants and phases in the processes are included [69]. In many cases, As-Is and To-Be flow-chart mappings are beneficial to use due to the visual and informative comparison between the current and future state.

In the recent decade, there has been a growing interest in utilizing new digital technologies, such as ML, in several sectors. The implementation often requires transformations in the work processes [76]. Digital transformations can be defined as the way digitalization changes the work processes. Digital transformations can affect the following levels [77]:

- *Process level*
- *Organization level*
- *Business domain level*
- *Society level*

This thesis will focus on the effects of digital transformation on the process level. Digitization is proven, in the literature, to affect several aspects of work processes, for instance planning, control, workers, flexibility, productivity, and quality [78–81].

In supply chain management (SCM), new technologies has given the improved potential to the decision-making support, changing the process by reducing the manual steps. In supply chain (SC) risk management predicting disruptions is of importance, and the technology can both help with predicting disruptions and support decision-making when it occurs. These decision-support models heavily rely on the completeness, fullness, validity, consistency, and timely availability of the data used. The integrated digital systems affect the ways of working by digitalizing the decision-making process [82].

There are several findings in the literature where digitization has been studied from a PM perspective. With the fourth industrial revolution, there have been changes in several sectors, including the manufacturing industry, food sector, agriculture sector, and energy sector [83–87]. Digitization in manufacturing has led to better asset efficiency, decreased waste of raw material, improved operational performance, and better logistics routing [80, 88]. In the construction sector, digitization can improve productivity and facilitate sustainable and safer construction [83]. In agriculture, new technologies can add value to the production by increasing the efficiency, and quality [85]. Digitization in the food sector has improved safety, reduced costs, increased quality, decreased process time, and assured quick response-time of the systems. Experiences of digitization in the food sector are relevant to look at when assessing digitization in the OR due to the similarities in the key performance indicators (KPIs).

New technologies have recently been successfully implemented in the food sector and led to changes in the manufacturing processes [84]. For instance, CV is used to automate and improve the analytical monitoring of the food by automatically detecting shape and color in real-time. This ensures that the food safety requirements are fulfilled and allow earlier detection of defects than the manual process, resulting in reduced food waste and recalls. The use of this technology solves the problem of human imprecision and variation in human judgment and contributes to cost reduction, customer satisfaction, increased efficiency, and improved food safety [84].

The effects of digital transformation seen in the food sector can be transferred to other domains, and one can expect similar outcomes. For instance, the success of the use of CV in the food sector can be transferable to the health sector. The transferability of the technology comes from the fact that large parts of ML-based methods are domain-independent [89]. However, there are uncertainties around success, such as data quality and integration with current work processes.

2.5   Preferences and Value Trade-Offs in the Health Sector

Medical errors are a matter of medical ethics, but the economic utility ethics also matters. Figure 2.8 shows a chart of the annual number of open and closed heart surgeries reported from the cardiac units in Norway from the year 1995 to 2020 [90]. For each year from 1995, there have been over three thousand cardiac surgeries, making up a significant expense for the hospitals in Norway. In an economic sense, it is incorrect to allocate resources to treatment without considering what benefits the resources could provide in alternative applications. In the last 25 years, the number of published empirical analyzes on how society should allocate resources optimally to evaluate the economy of the health measures has increased exponentially [91]. Kristiansen et al. [92] stated a goal for diagnosis and treatments to maximize patients' expected welfare and the society's life expectancy. This goal has some constraints concerning budgets and equality ideals since prioritization is made based on age, family situation, and type of disease [93]. It is relevant to assess the consequences of suboptimal diagnostics or treatment, both financially and in health. A calculation is done in Norway of expected health loss for common diseases [94, 95]. The health loss is lost years with good health quality.
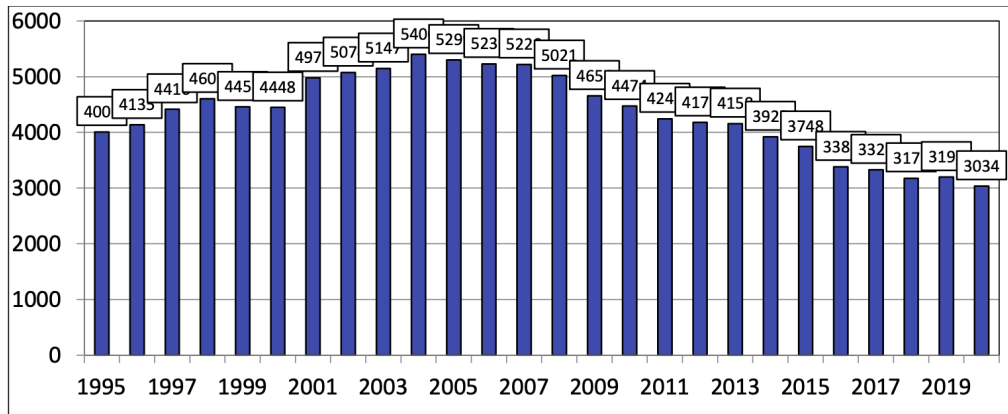
Figure 2.8: Annual number of open and closed heart surgeries reported from the cardiac units in Norway 1995-2020 [90]. Transcatheter aortic valve implantation procedures are reported and included until 2017. Intra-Aortic Balloon Pump Therapy treatments, implantation of pacemakers, and automatic implantable cardioverter-defibrillator are not included.

2.6 Machine Learning Fundamentals

ML is a subfield of artificial intellegence (AI). This research field studies computer-based algorithms that can complete tasks without being explicitly programmed. ML models have proven great performance when solving tasks related to predictions, classifications, or regressions due to the models' ability to improve performance through experience [96, 97]. ML models are applicable in a wide variety of domains, for instance, weather prediction, personalized advertisement, and risk management. The models are based on results and concepts from different disciplines of science, such as optimization theory, information theory, statistics, cognitive science, and control theory [98]. To design an ML approach, one has to choose the training experience type, a target function to be learned along with a representation of this function, and an algorithm that can use the training examples to learn the target function [97].

2.6.1 Artificial Neural Networks

Artificial neural networks (ANNs) define a group of learning systems inspired by the way the human biological brain behaves when solving problems; hence the name [99]. The physical neural network is composed of massively connected nerve cells, *neurons*, that process information. The information transmitted through the biological neural network is broken down, captured, and distributed in parallel in the interconnected neurons, so the information passed through each neuron is small [100]. The input to the ANNs are the initial data. The input data can for instance be images of animals that is to be classified as "dog" or "not dog". The output of the ANNs are the produced results of the given input.

The ANNs are composed of nodes with resemblance to the biological neurons, called *artificial neurons* [100]. The network is built up by nodes arranged in layers, and each node fires when a linear combination of the inputs exceeds some threshold [101]. Figure 2.9 shows a

simple mathematical model of an artificial neuron [100]. These artificial neurons, or nodes, are connected with *edges*, interconnecting node i with the successive node j. The strength of the connection is determined by each edge´s associated numeric *weight*, $w_{i,j}$. Each edge also propagates the output from node i to node j, corresponding to the *activation a_i*. After a node receives the activations a_i from predecessor nodes, a weighted sum of all the inputs is computed [101]:

$$in_j = (\sum_{i=0}^{n} w_{i,j} a_i) \tag{2.2}$$

Then, an activation function g is applied to the sum in_j:

$$a_j = g(in_j) = g(\sum_{i=0}^{n} w_{i,j} a_i) \tag{2.3}$$

The activation function g is a nonlinear function and is typically a logistic- or threshold function. This function makes the network able to represent nonlinear functions. [101]
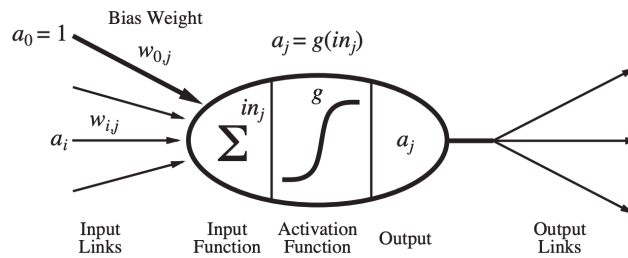
Figure 2.9: A simple mathematical model of a neuron [101].

ANNs consists of several connected nodes and can compute one or more outputs. The property of the nodes determine the performance of the computation and the network topology [101]. ANNs are composed of artificial neurons as nodes and weighted edges interconnecting the nodes. If the network architecture is a directed acyclic graph it is called a *feed-forward network*. Feed-foreward ANNs do not have an internal state except the numeric value of the weights.

2.6.2   Deep Learning

DL is a subfield of ML which represents complex models as a nested hierarchy of simpler non-linear estimators [102]. The DL model's architecture is an ANN with at least three layers, illustrated in figure 2.10. Each node in figure 2.10 is an artificial neuron as shown in figure 2.9. The DL models differ from most learning techniques by having a deep architecture, where the term *deep* represents the number of layers of estimators in the hierarchy [103]. Each layer in the hierarchy breaks the desired complicated task into simpler tasks. The input to the model is called the *visible layer* due to observable variables in the input data. The next layers have non-observable values; they are not given from the input data and are called the *hidden layers*.

The hidden layers extract increasingly abstract features, from the visible layer, to solve the task and return the output.

Deep neural network

Input layer   Multiple hidden layers   Output layer
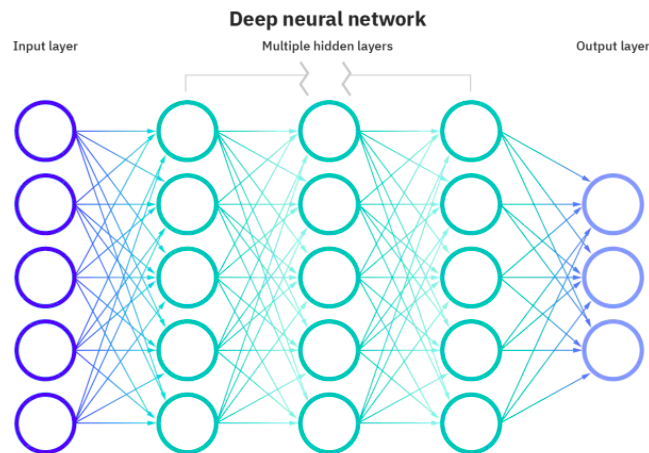
Figure 2.10: A simple deep neural network with an input layer, three hidden layers and an output layer. Each node in the network is an artificial neuron. The edges are interconnecting the nodes, and each edge has a weight associated with it [104].

2.6.3   CNNs

The idea of using ANNs for processing visual information resulted in the architecture named convolutional neural network (CNN) [105]. During the last ten years, CNN has given promising results in problem-solving related to pattern recognition, e.g., image classification, CV, and natural language processing [106]. The input to CNN models can be an image or a video. The structure of a CNN consists of convolutional layers, pooling layers, activation functions, and fully connected layers [107].

Convolutional layers

The convolutional layers aim to extract high-level features from the input image, by detecting patterns in neighboring pixels [108]. Each layer consists of a set of learnable filters, called kernels, which activates when detecting certain features in the input, for instance, edge detection [109]. A kernel is a set of weights in the shape of the receptive field, a spatially local section of pixels in the image to be convolved. The kernel has the same depth as the input image and is smaller in size [108]. A convolved image produce a feature map that indicates the probable locations of a detected feature in an input.

The kernel moves across every spatial location in the input until it parses the entire width

and height of the input, repeating the convolution process and creating the output feature map. The spatial increment of the kernel is called the *stride*. Increasing the stride can reduce computed data, and the models' sensitivity to spatial translation [108].

The convolution computes a sum of products between pixel values within the receptive field and the kernel weights. An image can be defined as the function $f(x, y)$, where x and y are spatial coordinates. The amplitude of f at any coordinate is called the *gray scale* of the image in that point [60]. The convolution $(w * f)(x, y)$ of an image $f(x, y)$ is shown in 2.4. w is the kernel of size m x n, where m and n are assumed to be odd integers. Here, $a = (m-1)/2$ and $b = (n-1)/2$.

$$(w * f)(x, y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s, t) f(x - s, y - t) \tag{2.4}$$

Pooling layers

A pooling layer is a downsampling process. The downsampling is done to look at the large features in the input. A pooling layer is commonly placed between the convolutional layers to achieve spatial invariance [109, 110]. Each feature map in the convolutional layer corresponds to a pooled feature map in the successive pooling layer. The pooling layers divide the input feature map into pooling windows of arbitrary sizes and compute one value from each window. This operation results in a pooled feature map with lower resolution. The most efficient method is max-pooling, which calculates the maximum value in the pooling window, as shown in 2.11 [110].
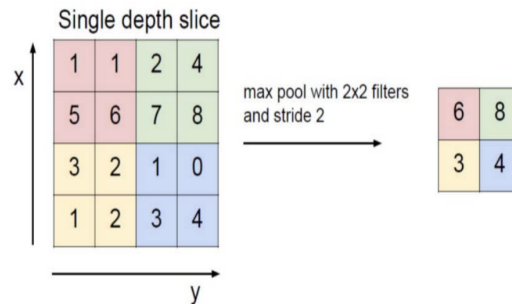
Figure 2.11: Illustration of max pooling with a kernel of size 2x2 and a stride of 2. [111]

Activation functions

The activation function affects the training dynamics and task performance. The function defines the output of the node given the weighted sum of the inputs. The most popular and successful activation function is the Rectified Linear Unit, defined as $f(x) = max(x, 0)$ [112].

Fully connected layers

The fully connected layers form the last layers in the CNN. It is a feed-forward neural network where every node is interconnected. The fully connected layers aim to learn a function from the feature space provided by the convolutional layers [113].

2.6.4 Visual Tracking

CV is a field that aims to extract information from images to solve tasks like object detection, segmentation, video classification, and object tracking [114, 115]. Most CV applications, such as cancer detection, face recognition, and automated driving, uses DL for efficiency [29, 116].

Visual tracking is a CV task, and the aim is to localize a given target in subsequent frames [49]. Due to its interdisciplinary application for real-world problems, visual object tracking is a growing research field which already is essential for medical image analysis [117, 118]. It can be challenging to track a real object from a video due to illumination changes, abrupt movements, or occlusion. Therefore an efficient and accurate tracking model is desired [119]. Siamese network [120] and discriminative correlation filter (DCF) [121] are popular tracking methods.

2.6.5 Transformer

The Transformer is a model architecture capturing global dependencies between input and output by utilizing self-attention mechanisms. The Transformer sat a new paradigm in machine translation, partly due to its short training time and its ability to parallelize tasks. It has an encoder-decoder structure. The encoder receives an input sequence of symbol representations and maps them to a sequence of continuous representation. The decoder takes previous output as the input and generates an output sequence of symbols. The encoder and the decoder are built up by a stack of N identical layers, and both contain at least one sub-layer with an attention mechanism [122].
The attention mechanism allows models to train faster and often increase the testing performance. The input to the attention layers is sequential data consisting of vectors of numbers. The attention mechanism aims to compute a linearly weighted sum of the vectors and train the model to learn the optimal weights $(\alpha_1, ..., \alpha_n)$. The attention function f computes the weights with a given query q and keys $(k_1, ..., k_n)$, as presented in equation 2.5.

$$\alpha_i = f(q, k_i). \tag{2.5}$$

The Transformer is famous for solving natural language processing (NLP) tasks, and has recently been applied to vision tasks like image generation and object detection [49].

2.6.6 TransformerTrack

The TransformerTrack architecture, as shown in figure 2.12, is a transformer-assisted tracking method. It is built up as a Siamise-like pipeline, with a CNN backbone for feature extraction,

the Transformer for exploiting temporal context, and a tracking model [29, 49]. The goal with TransformerTrack is to track an object in individual video frames [49]. The response is a tracking response map, with confidences of where the tracked object is.

Backbone -TransformerTrack

The proposed model, TransformerTrack, uses a ResNet-50 model backbone for feature extraction [123]. An additional convolutional layer is added between the backbone and the Transformer, to reduce the number of channels for visual tracking [49].

Transformer -TransformerTrack

Wang *et al.* [49] applied the Transformer [122] in visual tracking. Essentially, there is less need for vision-specific inductive bias in transformer-based models, which makes it advantageous to apply to CV tasks [124]. Temporal information among successive frames is merely used in basic tracking methods. Wang et al. [49] introduced the Transformer to the tracking architecture in TransformerTrack to increase the tracking performance by looking at the frame-wise relationship in the input-frames. The transformation characteristics in a modified Transformer architecture are well suited for propagating context in a temporal domain.

The Transformer is composed of an encoder and a decoder. The parts are separated into two parallel branches in a Siamese-like tracking pipeline. The encoder in the top branch receives template patches, the pictures showing what the model should track, and make compact target representations to generate high-quality encoded features. The transformer decoder is in the bottom branch and receives the search patch and the previous template contents from the top branch. The transformer decoder facilitates the object search by various target representations and spacial indicators and generates a high-quality search feature [49].

In TransformerTrack the Transformer improves the performance of the tracking by suppressing the confidence of the distracting objects [49]. The model architecture of TransformerTrack is illustrated in figure 2.13.
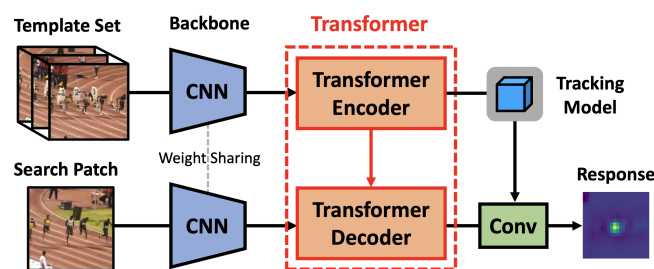
Figure 2.12: Model architecture of TransformerTrack [49].

Trackers -TransformerTrack

The TransformerTrack method connect a modified Transformer architecture with two different popular tracking baselines; the Siamese tracker [125] and DCF [126]. Both trackers, named

TrSiam and TrDiMP respectively, show improved performance and set state-of-the-art records on commonly used tracking datasets [49].

| Search Region | w/o Transformer | w/ Transformer |
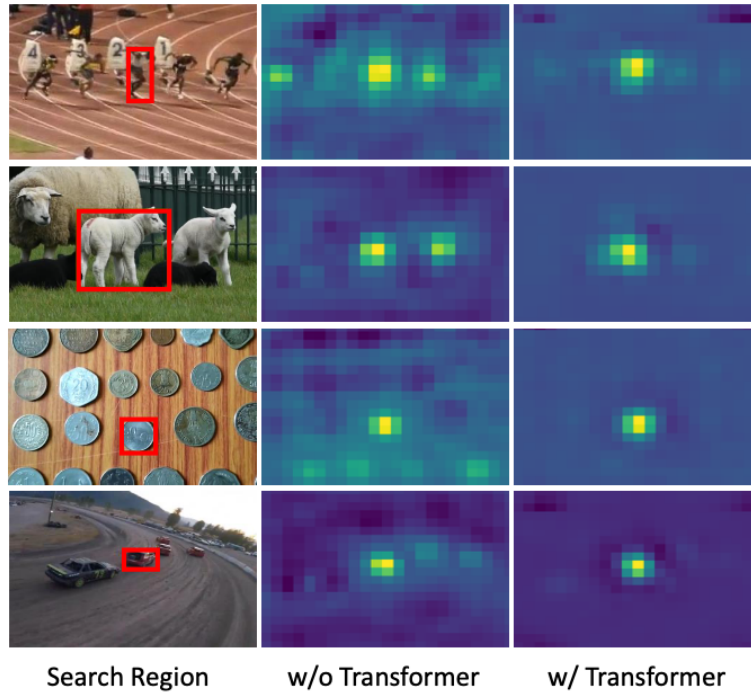
Figure 2.13: The first column shows the search region, with annotated targets, the second column shows the tracking response map of the DiMP baseline, and the third column shows the tracking response map of the TrDiMP [49]. The brighter the color, the higher the confidence.

2.6.7 Training ML models

The ML models need to be trained to solve different tasks. Most ML models, like Transformer-Track, apply to a wide variety of domains. When using the methods in a new environment, it must be tuned and tweaked to perform well. For example, the optimal parameters in TransformerTrack when tracking a car on the road are not the same as those for cardiac tracking purposes. There are no correct answers for tweaking the models.

Data preprocessing is necessarily to achieve optimal model performance. ML and CV are data-driven methods, and accuracy and performance largely depend on the quality of the input data. In addition to striving for the perfect combination of hyperparameters, a developer wants to present the relevant data in the best possible way. There are many processing techniques, and their success depends on the data, the model, and problem. If there is a lack of data, data augmentation is a technique to increase the amount. Data augmentation is also a way to increase the ability to generalize by introducing more diversified data. Examples of data augmentation techniques are rotation and flipping of the images.

The input data to the ML models are initially split into three different datasets. The model

train on the sample of the data called the *training dataset*. The *validation dataset* is used during training to provide an unbiased evaluation of the performance of the model *during* training. The validation data is previously unseen data. Lastly, the *test data* is the data used for testing purposes to evaluate how well the trained model performs on unseen data [127]. There are several split ratios, and the optimal split is dependent on the model and the data.

A ML model must be trained on data to perform tasks such as tracking and classification. There are three different learning paradigms in ML; supervised learning, unsupervised learning, and reinforcement learning, illustrated in 2.14. Before a supervised ML model can train on data, it has to be annotated [128]. Image annotation is the action of labeling images for ML model training purposes. The goal of the annotation is to ensure that the ML model learns the correct information [129]. For instance, the labels for tracking purposes specifies the tracking target, and the label is called the ground truth [130]. Bounding box annotation is illustrated to the left in figure 2.13, the ground truth is illustrated with a red bounding box [130]. The annotation can be done manually or automatically, depending on the problem complexity [130]. Human-annotated data often has a higher degree of accuracy [131]. In unsupervised learning, the main goal is to discover underlying patterns in the input measures, and the input data is not labeled. In reinforcement learning, there is no dataset to learn from and the model is learning based on experience and feedback on the model's behaviour [129, 132].
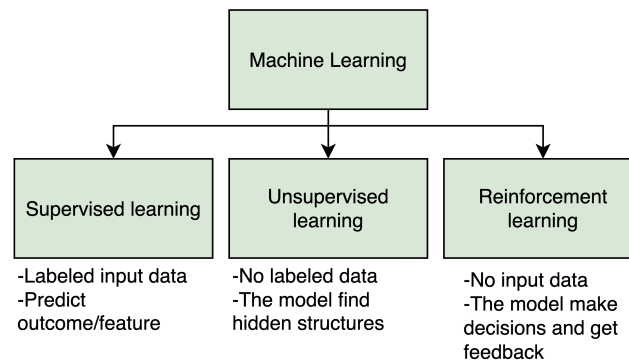
Figure 2.14: An overview of three machine learning paradigms with their main properties

Transfer learning is a method used to improve the supervised learning of ML models. Essentially, the knowledge from a source domain is leveraged for the target domain learning [133, 134]. The domains do not need to correlate, but the ML tasks must match. Figure 2.15 shows a simple architecture example of how the transfer learning method technically works. As illustrated in the figure, a part of a pre-trained method can be incorporated in the architecture of the target domain method. Transfer learning is often used in cases where the source domain has a vast amount of labeled training data and the target domain has a lot of unlabeled training data [133].

The network of a ML model is trained by minimizing the difference between the predictive output of the model and the ground truth. In the TransformerTrack the ground truth is a bounding
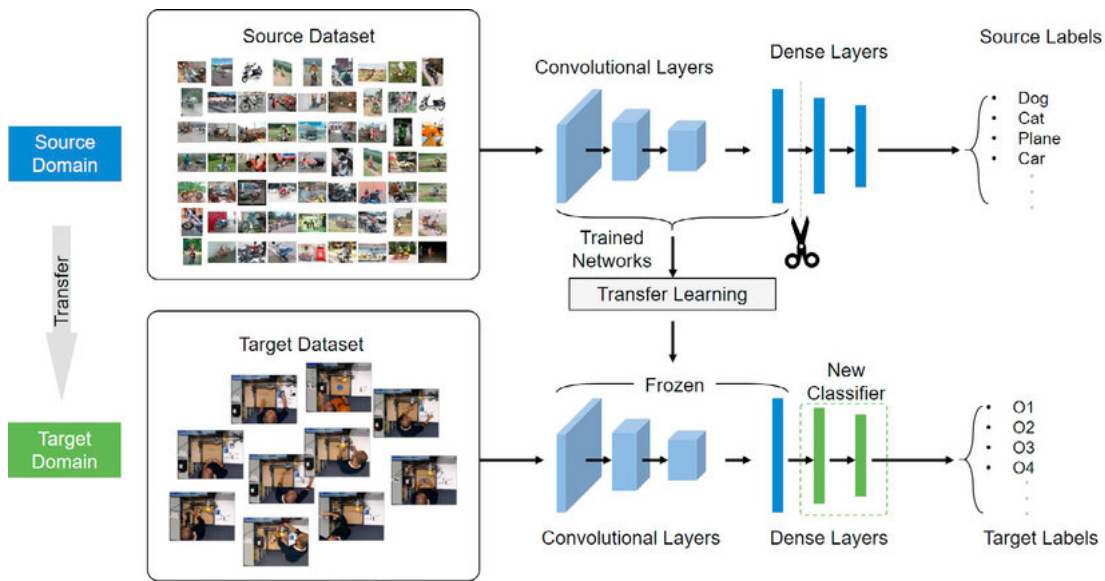
Figure 2.15: A simple architecture example of the transfer learning method [135].

box covering the object to be tracked, and the predictive output is the models tracking prediction of the bounding box. The difference between the ground truth and the predictive output is measured by a function computing the predictive error, called the loss function [136]. The loss function is used for quantitative evaluation of model performance. There exists several different kinds of loss functions, suited for different kinds of problems. TransformerTrack uses the Kullback–Leibler (KL) divergence loss function [137]. The KL loss function is applied in many aspects where it is desirable to quantify the difference between probability distributions, and is often used in speech recognition tasks [138]. For TransformerTrack it is feasible to use the KL loss function due to the response map, illustrated in figure 2.12.

Training loss is the loss computed on the training dataset, and validation loss is computed on the validation dataset. The training loss indicates how well the model performs on the data it is trained on, and validation loss indicates how well the model performs on new data [139] [140]. If the validation loss is high, the model performs poorly on new data, meaning it is not good at generalizing. If the training loss is high, the model performs poorly on the training data. Figure 2.16 illustrates the relation of the training and validation loss. Good performance on training data and poor generalization means that the model is underfitting, and poor performance of the training data and poor generalization implies that the model is overfitting. It is optimal to reach a balance between overfitting and underfitting to optimize the model [141]. The number of epochs decides how many times the network weights will be updated, and the number of epochs is a hyperparameter. An epoch means training the neural network with all the training data for one cycle, and the network weights are updated each epoch. The learning rate is another hyperparameter, and it determines to what extent the weights should be modified based on the calculated loss.
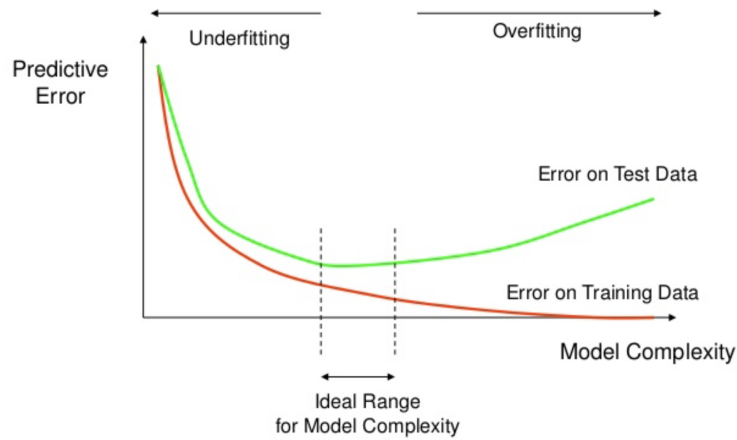
Figure 2.16: Overfitting and underfitting. The x-axis values are the epochs, and the complexity of the model. The value of the y-axis is the loss, the predictive error. The model is overfitting when the validation loss and training loss is low. The model is underfitting when the validation and training losses are high. The ideal range for model complexity is between the underfitting and overfitting area [136].

2.7 Digital Transformation with the Use of Machine Learning for the Cardiac Domain

Intelligent methods play an increasingly important role in the medical environment, and the systems are usually utilized for decision support [142]. Consequently, the nature of health care work is currently changing [143]. A CNN has outperformed or given as least as accurate findings as radiologists when detecting abnormalities on radiographs [144]. Some studies suggests that automated solutions with AI could help to detect heart failure [51]. Diagnostic imaging, especially ultrasound, is an essential asset in medical care [145]. Several measurements of cardiac function are based on anatomical information provided by the ultrasound images; thus, ultrasound data is used in various clinical applications [145]. In the last decade, DL has been applied to various medical ultrasound analysis tasks including cardiac-related tasks [10, 51, 146–148].

In section 1.1 the gaps in the current echocardiographic monitoring of the cardiac function were presented. The APOC-project is developing an ML-based tool to solve current problems with today's automatic monitoring methods. An automated cardiac monitoring system should be easy to use for physicians, provide measurements of relevant variables, be operator-independent, have a rapid response time, cause no harm to the patient, and be cost-effective [12]. ML has properties that can make it possible to develop a solution that meets all these requirements. Previous efforts to incorporate decision support models for clinicians has shown trouble with poor user interface, and the users struggle with a lack of time to use it [149, 150]. However, several surveys have concluded that the use of ML for decision support systems can be more helpful [151, 152].

Patient data acquisition is required to use ML models in the cardiac domain. DL models utilize vast amounts of data for training, often causing the data retrieval in a health care setting to be problematic [153]. The data retrieval in the medical domain requires approval from the patients and an ethics committee, which may be time-consuming. A skewed distribution of health conditions in the population of patients in the training dataset can introduce biases into the model [153].

It is advantageous that CV using DL can use ultrasound images obtained by traditional echocardiography [10]. The requirement for quality input data affects some doctors' monitoring choices. Several researchers performing myocardial strain measuring methods choose to estimate the global longitudinal strain due to the superior image resolution in the axial plane [20].

The echocardiographer can tune parameters in the ultrasound machine [154]. The quality and frame rate of ultrasound images affect the performance of tracking and strain assessment to a large extent [20, 155, 156]. A high frame rate results in a high temporal resolution; this means that the ultrasound system has an excellent ability to distinguish between instances in rapidly moving structures. The spatial resolution determines the quality of the ultrasound frames. There is a trade-off between frame-rate and spatial resolution in tracking performance because an increased frame rate will decrease the spatial resolution.

One of the most significant complications with implementing ML models into new disciplines is the fact that ML is non-explainable [157]. ML models receive input data with ground truths, then the model bases the output on relationships between input and ground truth that it has constructed on its own. Therefore, it is problematic for humans to comprehend how the model makes decisions. When applying ML models in a domain with small safety margins, like medicine, the difficulty with non-explainability might increase. Tonekaboni et. al [44] states that incorporating ML-based models effectively to clinical workflow requires establishing clinicians' trust. The need for understanding ML model behavior has led to the research on ML algorithms with increased explainability [158].

Another challenge with ML that is general for its application is that the methods are data-driven, which means that the models learn solely based on data [159]. It requires vast amounts of data for complex models to perform well. Unclean and noisy data result in poor model performance, and the ultrasound data has a tendency to be obscured [160]. The development of ML-based tools may be delayed due to awaiting patient approval and external ethics and privacy committees. To continuously improve CV models after it is deployed in production, it is usual with continuous data acquisition for further learning [161]. For example, the ML algorithms behind self-driving cars train on new images of the road while in use to continue to improve it [162]. If this technique is applied to a cardiac surgery setting, getting the approval of each patient undergoing surgery will be comprehensive.

Although the use of ML-based tools has proven to improve the performance of medical diagnosis, the tools can also decrease the accuracy of the doctors [163, 164]. When the technology

is implemented in a clinical workflow, the doctors are prone to over-reliance on the support, which might lead to decreased performance of the clinicians if the model performs poorly and make incorrect decisions [163, 165]. Grote et al. [163] suggest that the clinicians and the ML models must collaborate on decision-making to avoid this problem.

There are several advantages to using ML over hardware-based medical devices in the OR [166]. It enables faster and more iterative design and development. Additionally, addressing malfunctions and improving performance based on user feedback is more efficient with software products.

Chapter 3

Data and Method

Chapter three presents the data used for visual object tracking and the applied methods. Data retrieval, data preprocessing, and visual object tracking was primarily implemented with Python. Figure 3.1 illustrates an overview of the main Python pipeline of the project.

3.1 Data

The performance of ML models is largely affected by the representation, and the amount of input data [102]. This section describes data acquisition, metadata, and preprocessing. Read more about the packages, interfaces, and libraries used to preprocess the data in appendix B.

3.1.1 Data Retrieval

St. Olavs Hospital provides the data. The data available in this study are 2D TEE 2-channel and 4-channel B-mode images. Additionally, 3D TEE B-mode images were utilized to increase the training dataset size. The medical images are video frames of the cardiac cycle in different patients, and the number of cardiac cycles in each video sequence varies. A cardiologist with expertise in echocardiography acquired the images during TEE examinations of patients with an existing cardiac problem. The approval from the ethics committee to use the images for research purposes was in place before the beginning of the study, and patient consent was obtained individually. All images were anonymized.

The available data consists of a set of annotated 2D images (referred to as dataset 1), a set of 2D images without annotation (referred to as dataset 3) and a set of 3D data (referred to as dataset 2). The 3D ultrasound recordings were converted to 2D images to suit the specific tracking task by extracting images rotated about the depth axis. The datasets with the TEE images and annotation are presented in table 3.1.
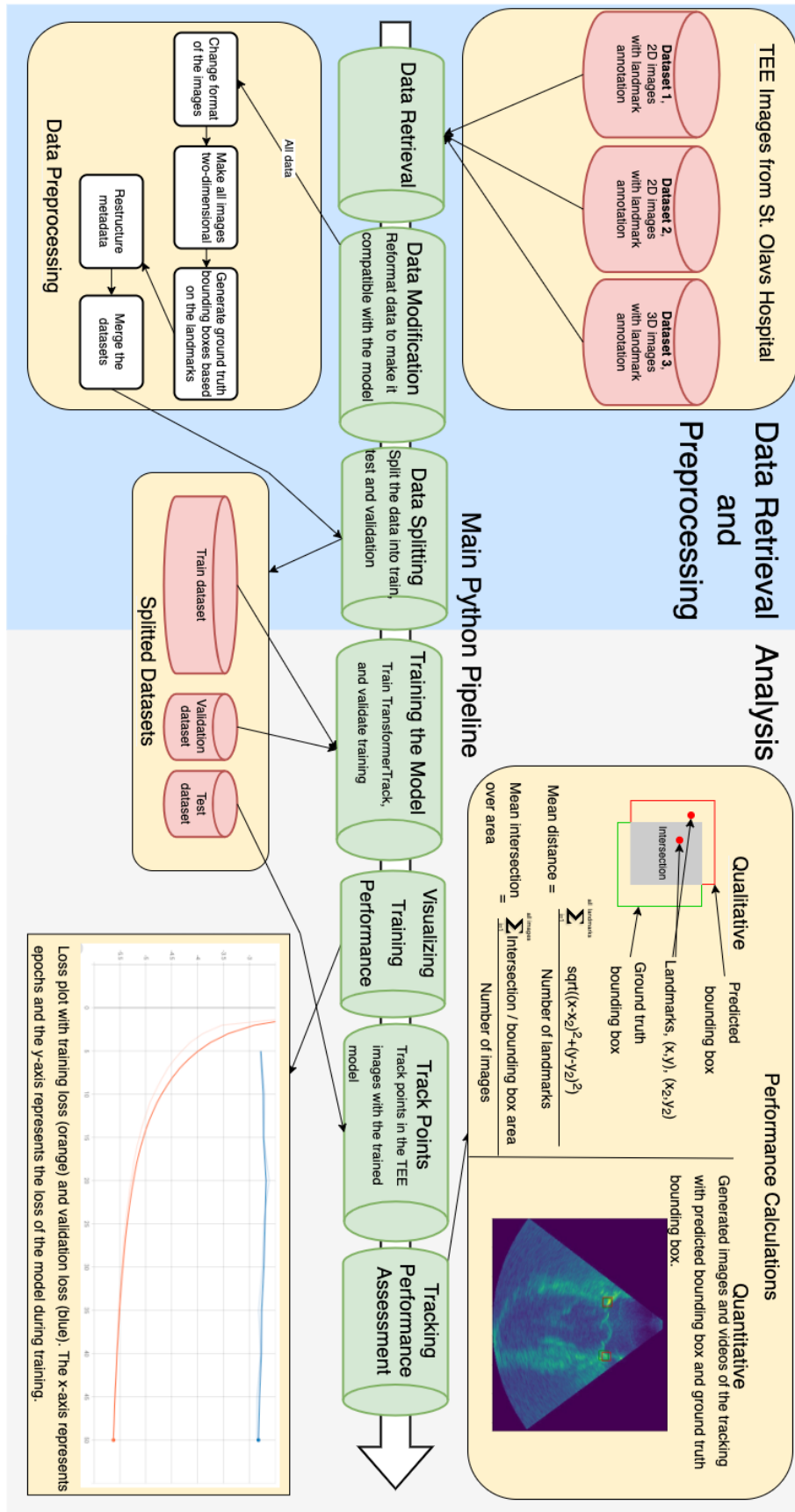
Figure 3.1: Main Python pipeline. The blue sections represent the steps of the data retrieval and preprocessing, and the grey section represents the key processes in the analysis.

Name	Original dimension	Annotation type	Annotated objects
Dataset 1	2D	Manual landmark annotation	Mitral points and myocardial points
Dataset 2	3D	Manual landmark annotation	Mitral points
Dataset 3	2D	Automatic landmark annotation	Mitral points
All data	2D and 3D	Manual and automatic	Mitral points

Table 3.1: Overview of the datasets containing the TEE images and annotation, provided by St. Olavs Hospital

3.1.2 Data Annotation

Landmark annotation is used. Figure 3.2 is an example of landmark annotation on a 2D TEE image. Each frame in the medical dataset has two landmarks on each side of the LV. In order to perform noise predictions, frames with poor visualization of the myocardium segment is annotated as noise, as shown in figure 3.3.
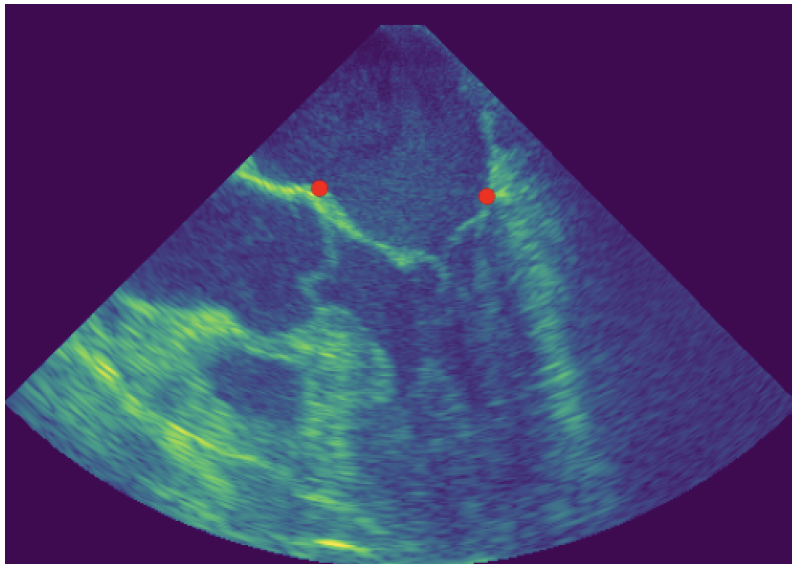
Figure 3.2: An example of data annotation on a 2D TEE image. The landmarks were placed on the mitral points.

The ground truth of each frame in the training and validation data corresponds to bound-
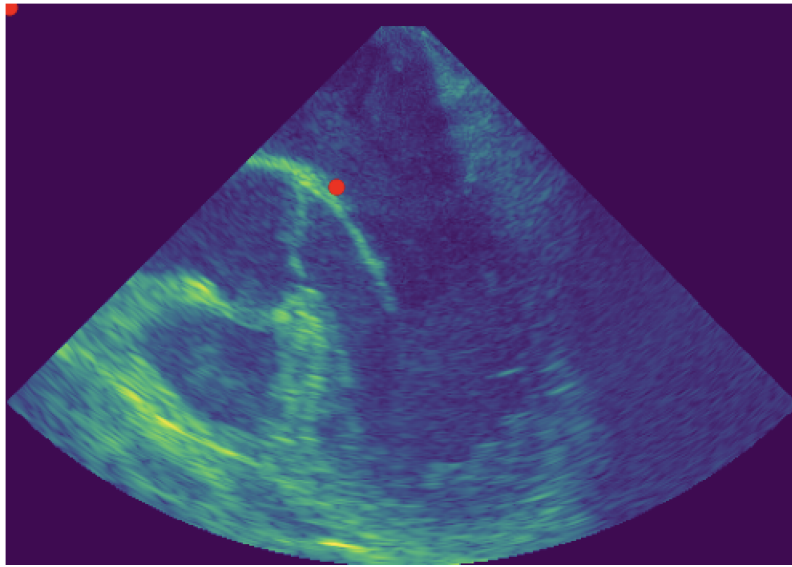
Figure 3.3: An example of manual landmark annotation on a 2D TEE image where one point is not found, and therefore marked as noise by placing the annotation in the upper left corner. The other landmark is placed on the left mitral point.

ing boxes. Since the medical datasets were annotated with landmarks, the bounding boxes are generated as illustrated in fig 3.4. The left landmark was placed in the upper right corner of the corresponding rectangular bounding box, while the right landmark was placed in the upper left corner. This placement is beneficial to the model's learning performance because the landmarks are placed on the edge of the myocardium, and then the bounding boxes will cover the most contrast around the landmark as possible. The goal of the TransformerTrack model is to predict the position of the bounding boxes in each frame.

Dataset 1 and 2 were manually annotated, and dataset three was automatically automated. The latter annotation is a semi-supervised approach used to reduce the time used for annotation. An automatic annotation may be less accurate than the manual; therefore, the author of this thesis manually verified the annotations of the dataset. The manual annotation of the mitral points in dataset 1 was performed by a previous master's student working with the APOC project. Additionally, dataset 1 was manually annotated by the author of this thesis for testing purposes, and the landmarks were placed on the myocardium along the LV on both sides. Anders Austlid Taskén did the manual annotation of dataset 2. The automatic annotation was done by a program that the APOC project group has developed.

3.1.3   Data Preprocessing

A data preprocessing pipeline was applied to the medical images. The medical data is required to be in an accepted format to be used for training of TransformerTrack, so the h5 files were converted to a format suitable for the method. Each h5 file contains multidimensional arrays representing the medical images and the locations of the annotations.

Bouding box of left landmark

W* 2/3

H

H* 2/3

W

Bouding box of right landmark

W* 1/3

H

H* 2/3

W
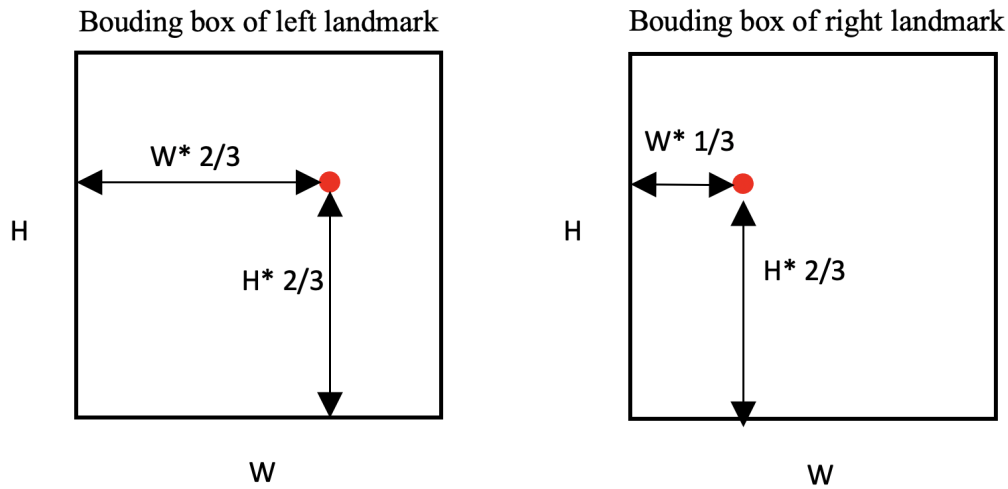
Figure 3.4: Illustration of the bounding box (width (W) and height (H)) in relation to the landmark (marked as a red dot).

The frames were first converted from h5 files to jpg files. The data were structured with a folder for each sequence, containing the frames, files containing the annotations and one metafile:

- Sequence folder

  - N images frames in jpg format
  - **absence.label** Nx1 array, shows binary whether an object is in each frame or not
  - **cover.label** Nx1 array, represents object visible ratios, ranging on a scale from 0 to 8
  - **cut_by_image.label** Nx1 array, indicates whether an object is cut by image in each frame
  - **groundtruth.txt** Nx4 matrix showing the bounding box: $[x_{min}, y_{min}, \text{width}, \text{height}]$
  - **meta_info.ini** shows meta information about the sequence (e.g. object class and movement class)

The medical data was split so that 80% of the dataset was used for training, 10% for validation, and 10% was used for testing. Figure 3.5 illustrates the required data preprocessing pipeline for the ultrasound data to be suitable for the TransformerTrack model.
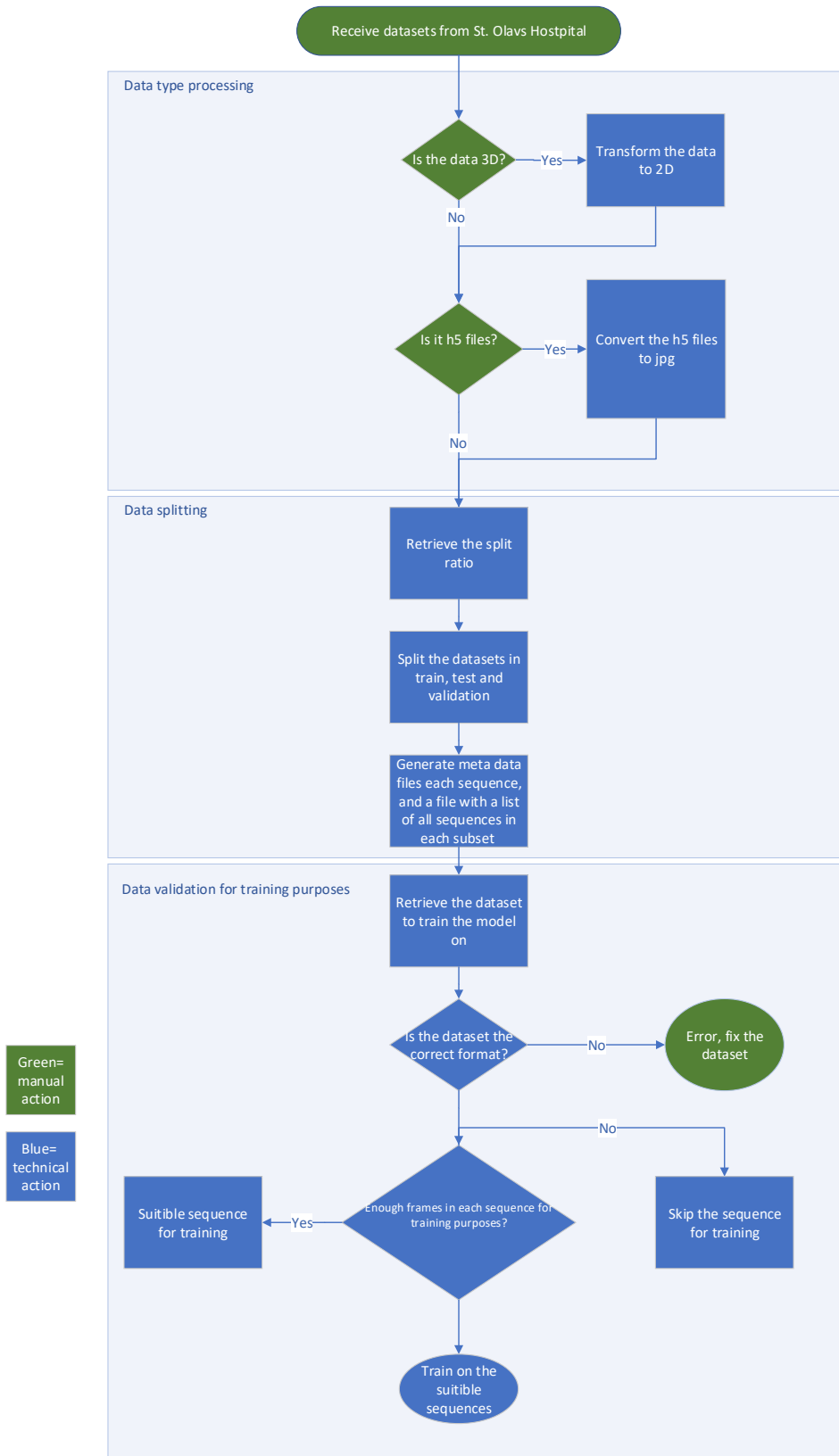
Figure 3.5: Overview of the required data preprocessing

3.2 Method

The method is composed of two parts; process analysis and landmark tracking using supervised DL. The literature review covers both parts.

3.2.1 Literature Review

A non-exhaustive systematic literature review was conducted, and the findings of the study are presented in chapter 2. The aim of the literature review is to gain an understanding of the theory of the research questions, as well as gain insight into digitization and PM, the state-of-the-art ML models, their area of application, and the implementation effect.

The literature search consisted of zooming steps due to the comprehensive scope of the resulting articles and books. The initial filtration was based on the title's relevance, the number of citations, and the year of publication. The year of publication was especially relevant to filtering articles on technologies since the aim was to research state-of-the-art methods. Figure 3.6 and figure 3.7 show that the number of hits on the query "machine learning health" has had a relatively rapid growth since the year 2000 compared to the search "cardiac anatomy", serving as an indicator of the popularity of that area of research. After narrowing the resulting hits down to about five articles per search and engine, the abstract was read. If the paper appeared relevant, it was skimmed and potentially read thoroughly. If an article or a book was highly relevant and of high quality regarding citation, spelling, appearance, and structure, the citation list was used to find other relevant articles. These zooming steps resulted in the papers or books in the bibliography list.
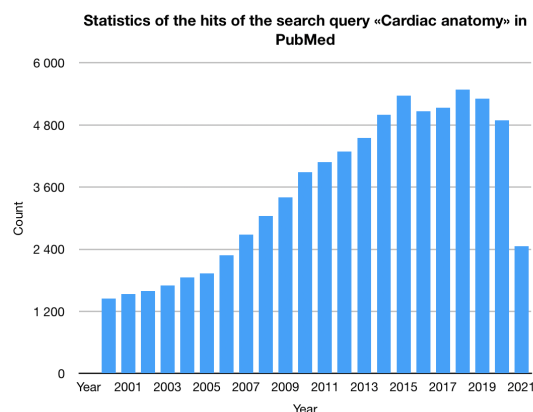
Figure 3.6: Statistics on the hits of the search query "Cardiac anatomy" in PubMed
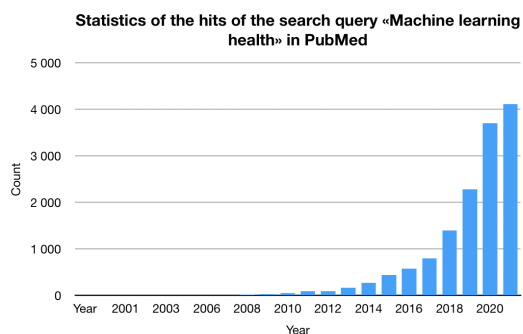
Figure 3.7: Statistics on the hits of the search query "Machine learning health" in PubMed

3.2.2 Process Analysis

Interviews were conducted with the primary goal of understanding today's work processes. The interviewees were doctors at St. Olavs Hospital. The questions were formulated to map

the nature of the different phases, the physicians' tasks during a surgery, tools in use, time use, decisions, and communication. These focus areas were based on the production management aspect in the OR, described in section 2.4. Another objective of the interviews was to confirm or disprove findings in the specialization project. Such clarification is significant for the correct analysis of the processes. Additionally, some questions were formulated based on the findings in the literature review presented in section 2.7. These questions were asked to investigate whether current problems or benefits of the use of ML in a health care setting apply to the monitoring tool.

The questions that the interviews were based on are presented in appendix D. The interview was not structural but rather a conversation. Follow-up questions were asked.

As-Is process analysis and To-Be analysis were conducted to evaluate the OR processes. The process mappings were based on the interviews.

3.2.3 Deep Learning

To automate the estimation of longitudinal strain with a DL based model on the TEE data, two points on the myocardium was located in each video frame of the cardiac cycle sequence. TransformerTrack was used for visual tracking of points in the heart, and the TrDiMP architecture was chosen. The tracking architecture is based on the work done by Wang et al. [49], and their transformer assisted tracking algorithm. The DL model was chosen since it has shown great performance on similar tracking tasks, and it has not yet been used in the medical domain. The specialization project preliminary study concluded that TransformerTrack might be feasible to use for tracking of TEE images. TransformerTrack sets several new state-of-the-art records and has given excellent results on the prevalent tracking benchmarks LaSOT [167], TrackingNet [168], GOT-10k [169], UAV123 [170], NfS [171], OTB-2015 [172], and VOT2018 [173] [49]. These state-of-the-art records show that the method has a good ability to track a wide variety of images.

Several experiments were performed to test the DL model. The model is designed to predict bounding boxes in each frame. The experiments tested the qualitative and quantitative performance on:

- Training of networks on all the datasets
- Tracking of mitral points
- Tracking of mitral points with different sizes of the bounding boxes
- Tracking of unseen points on the myocardium

To measure the performance of the predictions, the ratio between the estimated bounding box position and the ground truth bounding box was calculated, referred to as the mean intersection over union (IoU). The mean distance between the estimated cardiac point and the ground truth is another performance measure applied in this study. The equations for these measures are shown in figure 3.1. Additionally, the percentage of points marked as noise is measured to evaluate noise prediction performance. An image with visualization of the bounding box prediction and reference is generated from each frame, and a video is generated based on the

frames for easier manual assessment.

The networks were trained on data with mitral points annotation, because the annotation already existed on the given datasets. Additionally, the mitral points were more straightforward to annotate and track due to the contrasts of the mitral valve in the ultrasound images.

The model was trained on all of the datasets illustrated in table 3.1. After the training, a visualization of the training-and validation loss was generated, as shown in figure 3.1. A manual assessment of the loss plot was performed. The network weights at the epoch with the lowest validation loss were saved and used for tracking. The epoch with the lowest training loss was chosen if the validation loss was relatively constant.

For testing, the performance was measured qualitatively and quantitatively, as illustrated in figure 3.1. The tracking of the mitral points and tracking of unseen points on the myocardium were performed. The test set of Dataset 1 was used for measuring the performance.

The technical commands for training and testing are described in Appendix A.

Chapter 4

Results

This chapter presents the results from the interview, the As-Is and To-Be analysis, and the visual object tracking.

4.1 Process analysis

One aspect of the process analysis is to understand the work processes and how CV can help the medical team during cardiac surgery. The analysis can indicate improvements and opportunities with this tool, which is helpful for maximum utilization of the tool. The other aspect of the analysis is identifying substantial and specific changes in decision-making during the perioperative phase. It is essential to understand what will change with the automated tool to identify risks and make demands on the development of the tool.

An understanding of the current state is required to analyze the effect of the adoption of digital technologies [77].This section will present the current state and propose a future state based on the findings from the interviews. Therefore, the information is based on the work processes at St. Olavs Hospital. The As-Is process mapping is illustrated in figure 4.1, the To-Be process mapping is shown in figure 4.2, and processes after the prototype phase of the tool are illustrated in figure 4.3. The future To-Be process map displays how the APOC project imagines the tool to function after the prototype is working without problems. Further on in the thesis, "To-Be mapping" refers to the To-Be mapping of the prototype, in figure 4.2.

Overview of the Tool

The tool will acquire TEE images with the ultrasound probe that is already in the patient's esophagus pre-surgery. Once the ultrasound probe is in place, it does not require an expert echocardiographer to obtain images. TEE images will be taken automatically and automatically passed into the trained ML networks of the tool. The ML tool then estimates the quantitative measures of cardiac function continuously in real-time. The model's output will be a continuous curve integrating global and local heart functions. The doctors decide on a threshold, and the tool raises an alarm if the output exceeds the threshold. If the output exceeds the threshold, the heart function deviates considerably from the standard that a specialist should
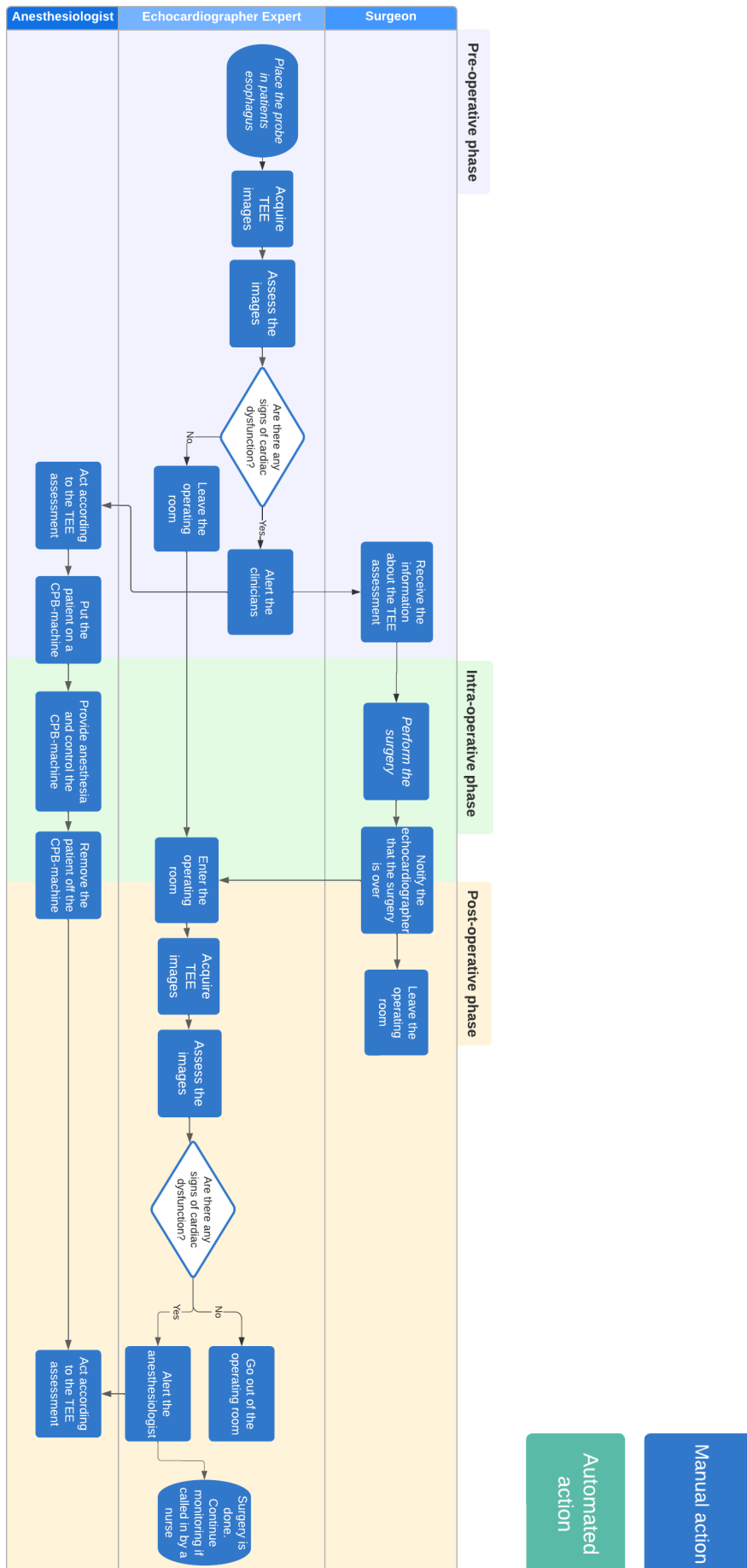
Figure 4.1: As-Is process map
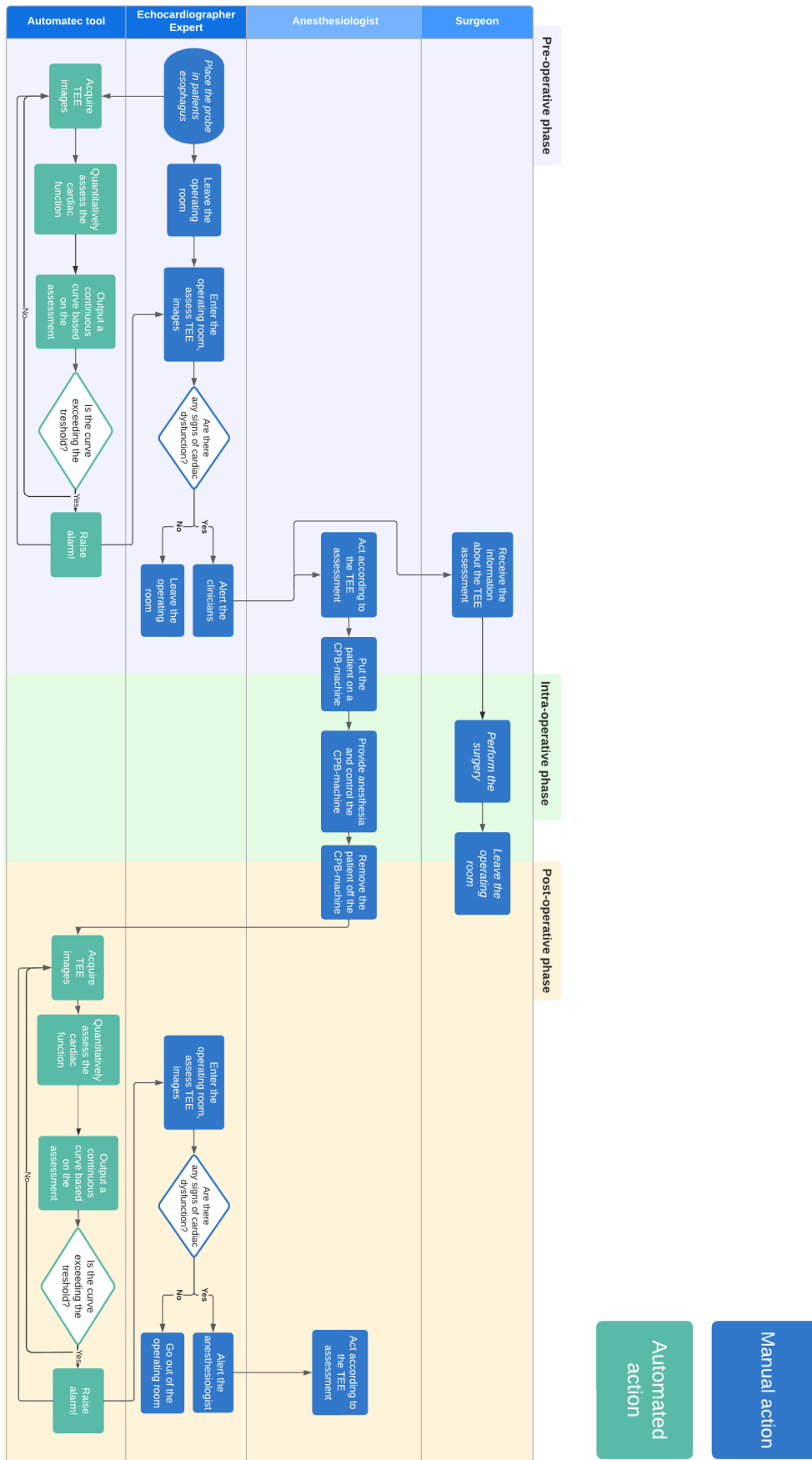
Figure 4.2: To-Be process map
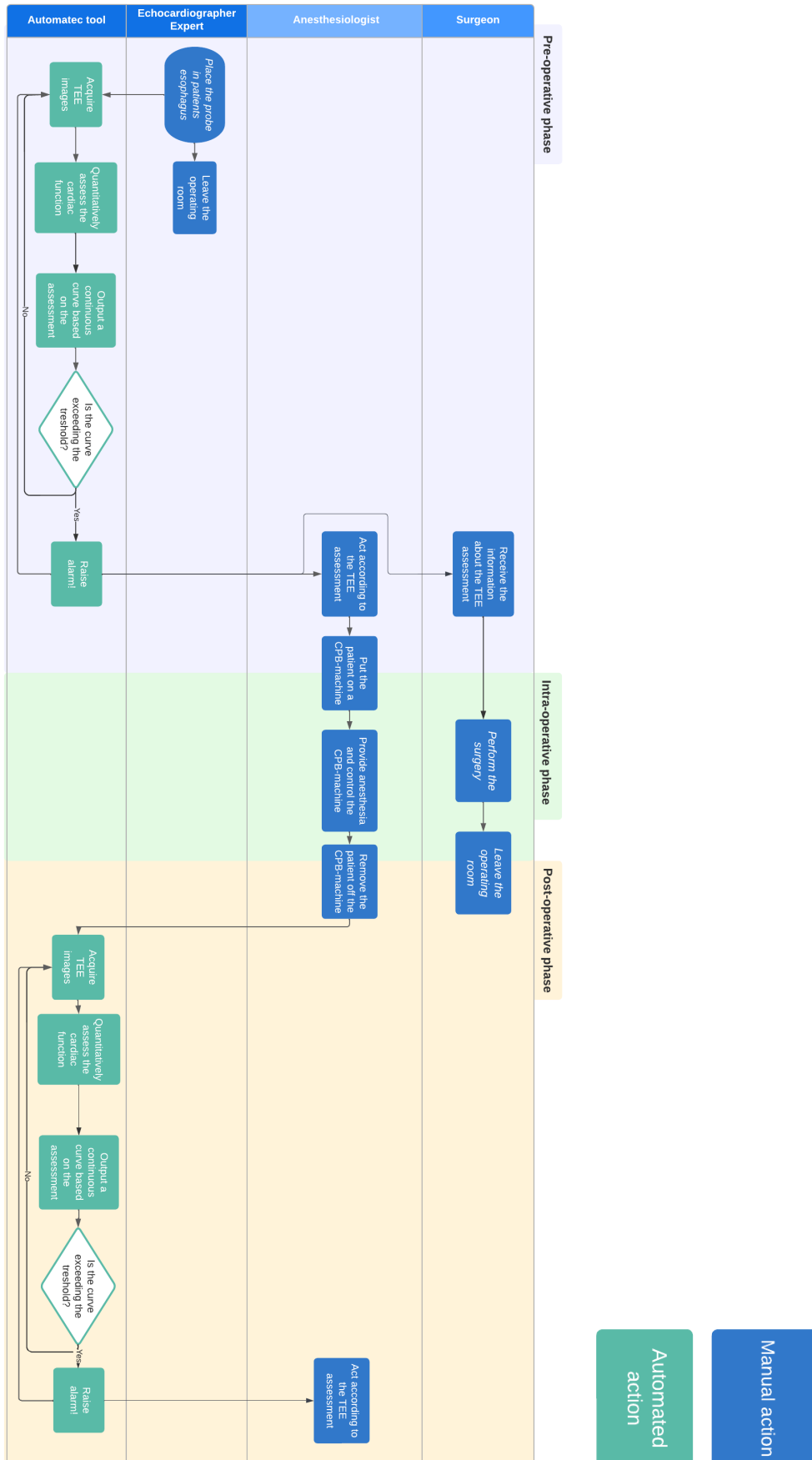
Figure 4.3: To-Be process map, in the future

be notified. This automatic monitoring process will be used during the perioperative phase of a cardiac surgery [4]. The tool will provide earlier detection of anomalies and facilitate preventative work in the OR. The goal is that the tool will complement the existing monitoring, thus increasing patient safety, decreasing time, and utilizing the information in the TEE images.

Main Stakeholders

The automatic monitoring tool will mainly affect three types of physicians: cardiac surgeons, thoracic anesthesiologists, and expert echocardiographers. The expert echocardiographer is either an anesthesiologist or a cardiologist, and in Trondheim, it is, in most cases, an anesthesiologist. Mainly, the expert echocardiographer will directly interact with the tool.
In a technical sense, the physicians do not find it challenging nor new to use ML in the OR, based on the testing of the prototype. The doctors do not differentiate between an old-fashioned automatic tool and an ML-based tool.

Processes in the Perioperative Phase

There are several stages of the perioperative phase of cardiac surgery. The As-Is process mapping illustrates the stages and the processes associated with the main stakeholders in figure 4.1. The flowchart is based on how the doctors perform surgery in Trondheim. There are often two cardiac surgeons, and their role together is represented in one swimlane in the mappings. The surgeon's task is to operate on the heart to fix the dysfunction. Doctors perform comprehensive examinations of the patient's cardiac condition before cardiac surgery. Therefore, the surgeon knows what the specific technical tasks are. The thoracic anesthesiologist is responsible for the patient's care, including giving anesthesia, adjusting medications, and controlling the CPB machine. The expert echocardiographer manages the ultrasound machine, acquires images, and assesses them. The anesthesiologist is the one who acts based on the findings the expert echocardiographer makes. Results from the visual assessment of TEE images rarely lead to an extensive change in the surgeon's procedure.

The echocardiographer acquires TEE images of the patient right before the surgeons initiate their procedure. These ultrasound images are used as a reference to see if the surgery affects cardiac function. While the surgeon operates, stopping the heart with a CPB machine is necessary. The automatic tool will not be used when the heart is not beating, nor will any traditional clinical variables. However, in the transitional phase between the machine and the natural heartbeat, the tool will actively provide an output on cardiac function. This is a critical phase of the surgery, with a high risk of complications. The transition is not abrupt since the blood is gradually sent back into the heart. Therefore, there is a need for all doctors present until the heart successfully starts beating on its own. Because of the nature of the procedure and the need for teamwork with all clinicians, the automated tool will not be remarkably beneficial in that phase.

The patient is sent to the intensive care unit in the post-operative phase when the surgeon has completed the surgery, where standard monitoring tools further monitor the patient's heart.

There are nurses present, but the doctors are not necessarily needed at this stage. The automated tool will, at that stage, serve an essential role. The tool raises the alarm if its output exceeds a given threshold, and the nurse calls the anesthesiologist. As the interviewed doctors stated, the tool will result in continuous monitoring rather than a discrete assessment.

Condition of the Hearts in the Dataset

Since the data is retrieved from patients undergoing TEE related to an investigation of an existing disease, there are most likely no completely healthy hearts. A model trained on a dataset solely consisting of patients with one type of cardiac function may have a poor ability to generalize. It is difficult for a developer with limited knowledge of cardiac anatomy to consider whether the dataset consists only of hearts with one type of disease. Therefore, the physicians were asked if it could be problematic for a model to be trained on this dataset and not on healthy hearts. The physicians agreed that none of the images in the dataset likely show healthy hearts. However, the ultrasound images in the datasets represent a vast panorama of conditions in the heart. Some have arrhythmias in the heart, and others have valve disease. Some have poor pumping function, while others have a heart that pumps more blood and contracts more strongly than a healthy person. The combination of conditions in the heart results in a dataset with all degrees of heart function. Since the goal of the model is to output a value on heart function, the physicians stated that the lack of a healthy population in the datasets would not affect the model's performance poorly.

Deciding the Threshold

The tool's output curve is a composition of the value of several cardiac functions, including the strain value. Some post-processing must be done based on the cardiac functions to generate the global value. As stated earlier, the clinical monitoring tool will function so that its output is a curve, and if the curve exceeds a threshold, an alarm is raised. If the continuous output curve of the tool exceeds the threshold and raises the alarm in the correct case, this is called true positive (TP). However, if the alarm is wrongly raised, this is false positive (FP). In the same manner, if the curve does not raise the alarm in the correct case, it is called true negative (TN), and if an alarm should have been raised, it is as false negative (FN) output. The possible outcomes of the monitoring tool is illustrated in 4.4

The process analysis in the specialization project stated that the model's threshold was predetermined and fixed for all patients. A fixed threshold value potentially leads to inaccuracy and an increased number of FP and FN. Additionally, it was stated that the developers implemented the threshold, and therefore extra direct communication was needed between doctors and technicians. However, this is not right, and in the interviews, the clinicians cleared up the misconception. The threshold is different for each patient and decided by the doctors. The value is adjusted during the pre-and post-operative phase if the patient's heart function changes.

If the threshold value is too low, the tool will erroneously raise the alarm more frequently, and an expert echocardiographer must needlessly show up. If the threshold is too high, more false negatives will jeopardize the patient's health.
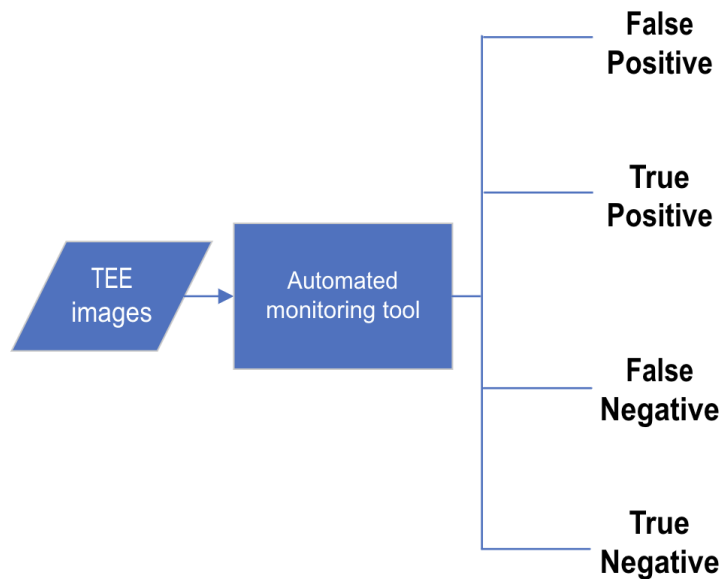
Figure 4.4: Illustration of the possible output outcomes of the monitoring tool.

As illustrated in the To-Be mapping, in figure 4.2, if the prototype tool raises the alarm, a clinician needs to check if it is correct. The doctors stated in the interview that they want to reduce the FP alarms to prevent them from checking unneeded. On the contrary, they said that the physicians must carefully decide the threshold for patient safety.

Communication

As illustrated in the As-Is and the To-Be process mappings, the tool's implementation will affect the communication between the staff. In the pre-and post-operative phase, the echocardiographer expert assesses the TEE images and shares the findings with the anesthesiologist and the surgeon. Since the echocardiographer expert validates the output of the tool during the testing of the prototype, the communication in the pre-and post-operative phase will not change. However, the future To-Be map shows that the monitoring tool will digitalize this communication. The tool will notify the physicians if it estimates cardiac dysfunction by raising the alarm. Because the monitoring tool is not in use in the intra-operative phase, the communication will there remain unchanged.

Workflow

The monitoring tool will affect the workflow in the pre-and post-operative phases. Today, the echocardiographer expert enters the OR before and after the surgeon's procedure to manually assess TEE images. Additionally, the nurse may summon the echocardiographer expert to the intensive care unit to assess the TEE images. The monitoring tool transforms the assessment of

the TEE images from discrete to continuous, thus changing the workflow. As illustrated in the To-Be map, the echocardiographer expert is called into the OR in the pre-operative phase or into the intensive care unit in the post-operative phase if the monitoring tool raises the alarm. A high FP ratio will result in unnecessarily high demand for the expertise of these specialists, which may result in higher costs and increased operational time. Therefore, there is a goal to remove the manual validation of the tool's output after a while. If the manual verification is removed too soon to reduce extra costs and time, it may affect patient safety.

The Benefits of Mapping Work Processes

A stakeholder performing an isolated code evaluation may not fully understand the tool's safety, potential risks, and effectiveness. The mapping of the tool's role in the OR may facilitate the evaluation of the tool for stakeholders. Additionally, it might prompt developers and users to equally understand the expected operational conditions. A better understanding of the tool's role in the perioperative phase can make it easier to make demands on ML development. The visualization of today's processes clarifies which personnel will use the tool, which may help the developers ensure that the tool's output is suited for the competence of all users.

The process mappings give the developers a clear insight into the various clinical work procedures. Since software as a medical device enables frequent and iterative maintenance and repair based on user feedback, it may be beneficial for developers to understand the physicians' tasks when improving the tool based on their feedback. The To-Be mapping can also be helpful when understanding and identifying potential risks.

Visualization of expected future processes can contribute to the transparency of the tool's objectives by making them easy to comprehend with merely a few glances. For example, it becomes clear for the nurse and the developer that the tool raises the alarm if the output curve exceeds the threshold and not the nurse. The As-Is and To-Be process mappings can also help communicate the benefit of the tool to third parties. Lastly, the mappings may remind the developers to integrate the user experience aspect in the development, thus giving extra thought to the nature of the tool's future environment.

4.2   Tracking Results

In this section, the results of the visual object tracking is presented in the form of experiments. The experiments provides the performance of training and testing on different datasets and with different parameters.

4.2.1   Experiment 1: Training Performance

The network was trained four times, once on each dataset. The name of the trained networks are corresponding with the name of the training dataset, shown in 3.1. All datasets were split with the same ratio, and all parameters were the same. All the networks were trained with 50

epochs. To read more about the datasets, see section 3.1.
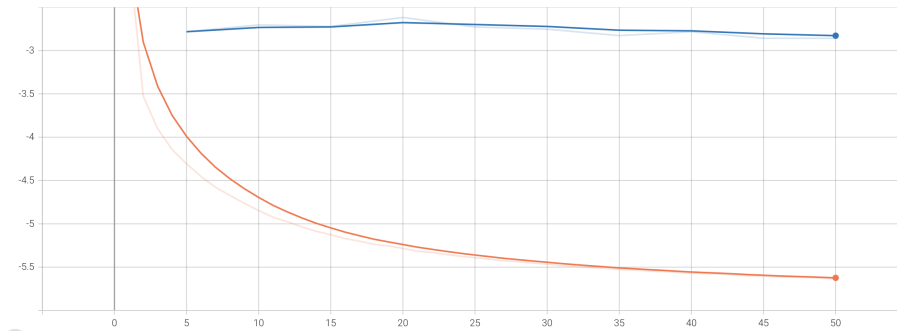
Figure 4.5: Training (orange) and validation (blue) loss of Network 1. The x-axis represents the epochs, and the y-axis represents the loss values.
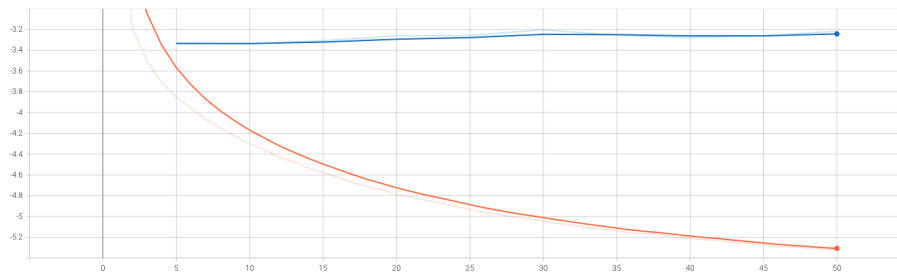
Figure 4.6: Training (orange) and validation (blue) loss of Network 2. The x-axis represents the epochs, and the y-axis represents the loss values.
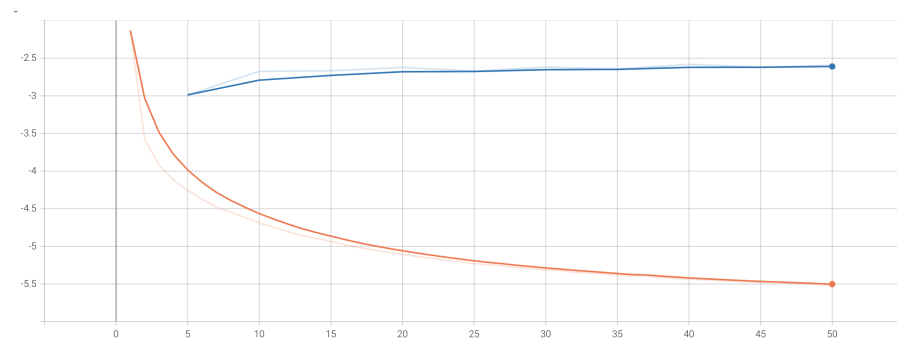
Figure 4.7: Training (orange) and validation (blue) loss of Network 3. The x-axis represents the epochs, and the y-axis represents the loss values.

The settings on the training performance plot were initially so that the validation loss was calculated from epoch 5. However, it was interesting to investigate if there existed any change in the validation loss before epoch 5. The result of a network trained on dataset 3 with five epochs is presented in 4.9.
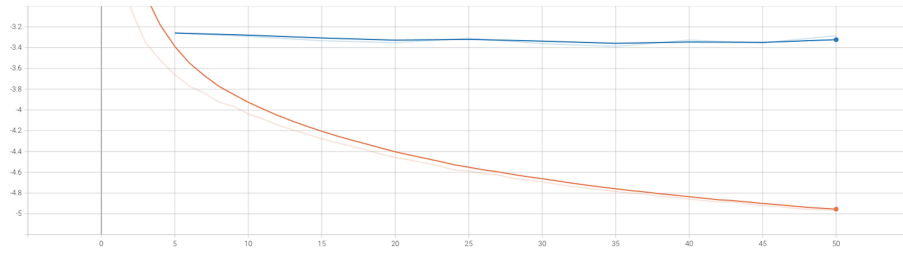
Figure 4.8: Training (orange) and validation (blue) loss of Network 4. The x-axis represents the epochs, and the y-axis represents the loss values.
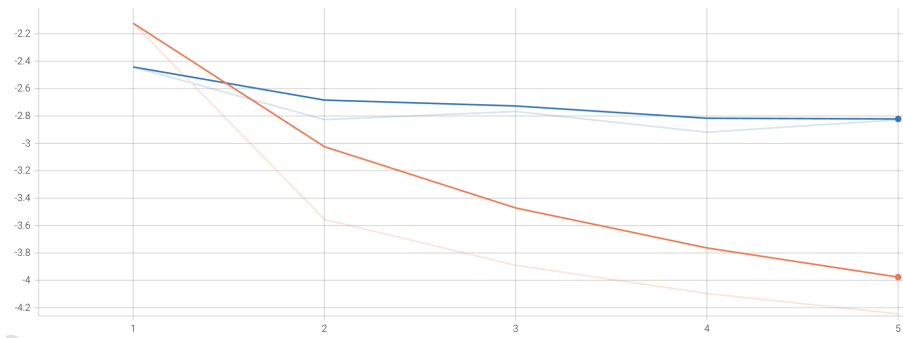
Figure 4.9: Training (orange) and validation (blue) loss of a network trained on dataset 3 with 5 epochs. The x-axis represents the epochs, and the y-axis represents the loss values.

4.2.2   Experiment 2: Tracking Performance

The first tracking experiment was to test each network, presented in the previous section, on the test set from Dataset 1. Network 5 is the network trained solely on the GOT-10k dataset with high-diversity videos of animals moving in the wild [169]. The models were tested on the same test set to have the same prerequisites for the performance measurements.

Table 4.1 presents the results of the quantitative performance measures. It is valuable to explore which dataset used for training provides the best-performing network. The best-performing networks have the highest mean intersection over union and lowest mean distance. The landmarks marked as noise do not say anything explicit about performance since it can be positive that the model does not track the landmarks in a blurry image, and it can be negative if the image quality is fine. If the percentage of noise is relatively high, the other measures are computed based on fewer data. Therefore, it is essential to consider the ratio of landmarks marked as noise when evaluating the other measures.

The qualitative image result of one frame from the tracking of the points at the mitral points landmarks at the myocardium is illustrated in 4.10.

The video result of tracking mitral points with Network 3 on a video sequence of one patient is shown in this link:
Press here to view the video results of tracking of mitral points

Network	Mean intersection over union	Mean distance in pixels	Standard deviation	Noise
Network 1	0.414	6.105	5.190	14.744%
Network 2	0.340	7.112	5.439	14.744%
Network 3	0.432	5.565	4.763	14.744%
Network 4	0.429	5.919	5.378	14.744%
Network 5	0.349	13.521	22.289	14.744%

Table 4.1: Quantitative measures of tracking performance. The networks are trained on different datasets, corresponding with the name of the datasets. Network 5 is the pre-trained network. All networks are trained with 50 epochs. All networks are tested on the same test dataset.
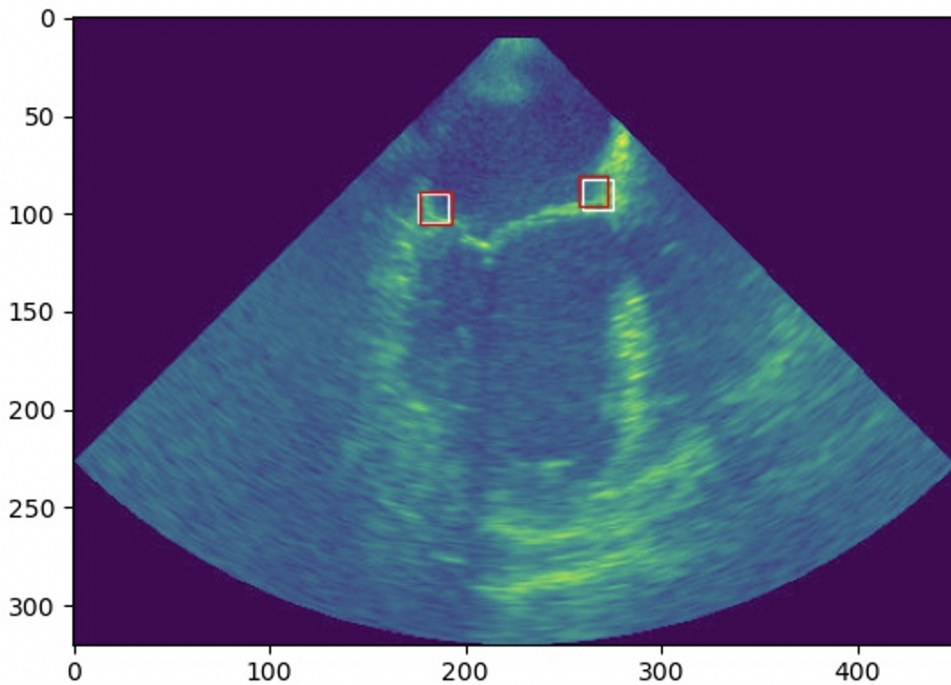
Figure 4.10: Qualitative output of one frame of the mitral points tracking. The red bounding box is the model prediction, and the white is the ground truth. The x-axis and y-axis values are pixels.

The red bounding box is the model prediction and the white is the ground truth.

4.2.3 Experiment 3: Bounding Boxes

Experiment three tests the training and tracking performance with varying sizes of bounding boxes. The bounding box size of the other experiments were 15 pixels. Dataset 3, that resulted in the best performing network in Experiment 2 based on the measures in table 4.1, were used

for training in this experiment.
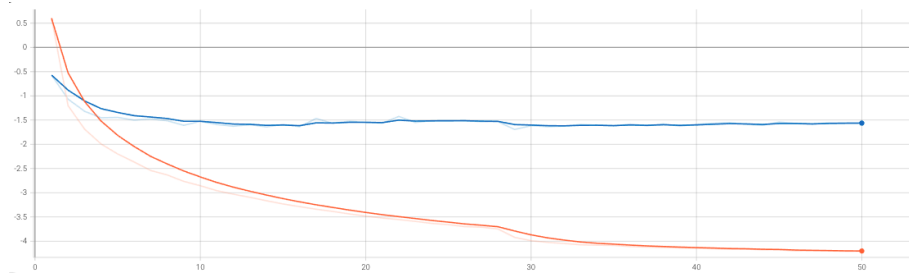
Figure 4.11: Training (orange) and validation (blue) loss of a network trained on Dataset 3 with a bounding box size 5. The x-axis represents the epochs, and the y-axis represents the loss values.
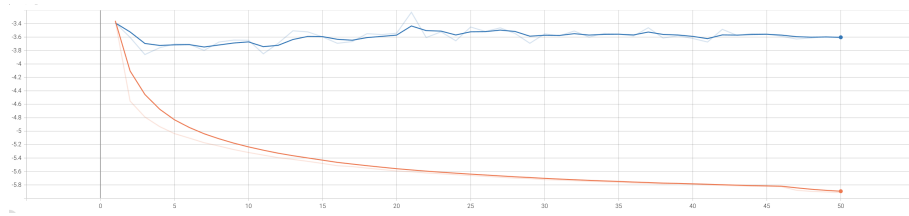
**Figure 4.12:** Training (orange) and validation (blue) loss of a network trained on Dataset 3 with a bounding box size 25. The x-axis represents the epochs, and the y-axis represents the loss values.

The loss plot in 4.12 shows that the training loss converge to -5.8 and the validation loss converge to -2.6. The loss plot in 4.11 shows that the training loss converge to -4 and the validation loss converge to -1.5. The loss plot in 4.7 shows that the training loss converge to -5.5 and the validation loss converge to -2.5.

The three networks trained on Dataset 3 with bounding box sizes of 5, 15, and 25 pixels were used to track the mitral points. The performance measures of the tracking are presented in table 4.2.

Bounding box size in pixels	Mean intersection over union	Mean distance in pixels	Standard deviation	Noise
5	0.297	8.903	5.840	39.940%
15	0.432	5.565	4.763	14.744%
25	0.355	9.446	17.675	39.940%

Table 4.2: Quantitative measures of tracking performance with different bounding box sizes. The networks are trained on Dataset 3 with 50 epochs. The mitral points are tracked.

4.2.4 Experiment 4: Track Unseen Points

For strain estimation, points on the myocardium were tracked with the networks trained on the mitral points. Figure 4.13 shows one frame with the ground truth bounding box on the myocardial points. The performance measures are presented in table 4.3.

Network	Mean intersection over union	Mean distance in pixels	Standard deviation	Noise
Network 3	0.234	11.692	10.172	39.940%

Table 4.3: Quantitative measures of tracking performance of unseen points. The "lower" myocardial points is tracked. Bounding box size is 15 pixels.

An example of the qualitative output of the tracking of the landmarks at the myocardium is illustrated in 4.13.
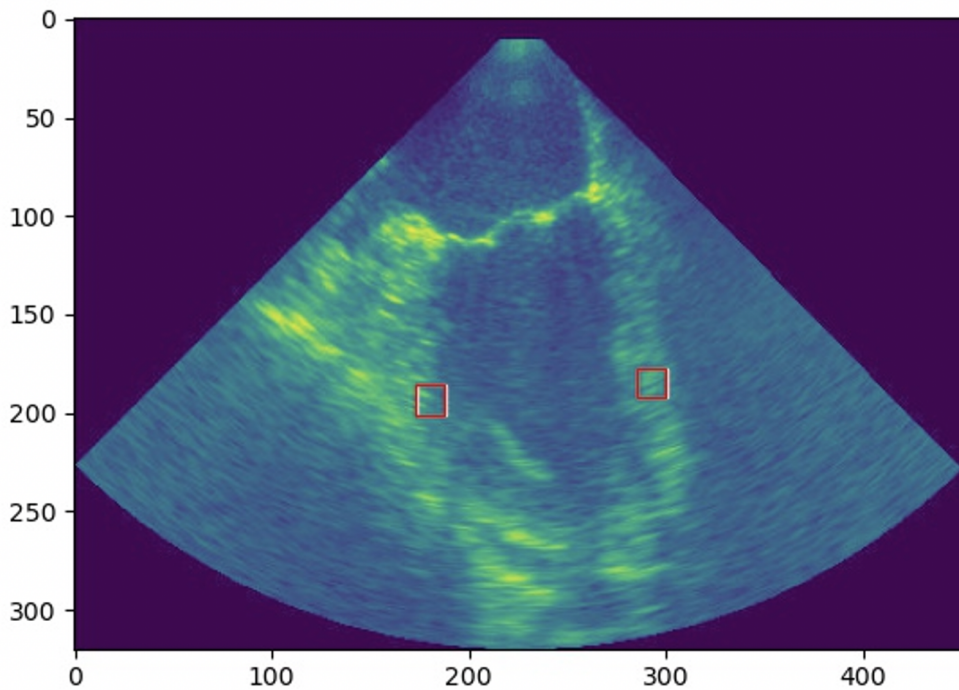
Figure 4.13: Qualitative output of one frame of the myocardium points tracking. The red bounding box is the model prediction and the white is the ground truth. The x-axis and y-axis values are pixels.

The video result of tracking myocardial points with Network 3 on a video sequence of one patient is shown here:
Press here to view the video results of tracking of myocardial points
The red bounding box is the model prediction and the white is the ground truth.

Chapter 5

Discussion and Further Work

This chapter will discuss the results presented in the previous chapter, provide a brief analysis of the results, present the limitations, and suggest further work.

5.1 Discussion

The discussion is divided into three parts. The two first sections belong to each research question. The last section provides an analysis of the work in this thesis.

5.1.1 RQ1: How will the monitoring tool affect the processes in the perioperative phase?

There are both advantages and disadvantages to integrating ML-based software in a clinical setting. From the process mappings, it appears that the monitoring tool will contribute to less inter-and intravariability in clinical decision-making by digitalizing the manual assessment of the TEE images. Additionally, an advantage of automated perioperative measures is reduced workload for cardiologists or anesthesiologists. However, a downside with the monitoring being automated is the possibility of physicians' over-reliance on the tool.

The tool will facilitate preventative work in the OR and improve the accuracy of the operational surgery and reduce the operating time. However, the safety can decrease due to a high FN rate and physicians' over-reliance on the tool. Operating time will increase if the output is FP and the tool wrongly raises the alarm. During the prototype phase of the tool, the expert echocardiographer manually automates the tool's output to identify FP outcomes. The manual double-check ensures that the anesthesiologist does not act based on wrongful information from the tool and contributes to maintaining patient safety. Patient safety might increase because the tool transforms the TEE-based monitoring from intermittent to continuous. Caridiac dysfunction can be detected earlier by constant monitoring, thus increasing patient safety and increasing the operational time.

ML technology seems suitable to assist with medical decision-making. Previous decision-support methods have suffered from poor user interface and physicians' need for extra time to use them.

The interviews showed that the physicians found the monitoring tools' incorporation into the OR satisfactory. The tool does not require much user interaction, and physicians are not required to acquire new technical knowledge. However, the echocardiographer expert needs to tune the parameters of the ultrasound machine, for example, the frame rate, to obtain images that provide the ML tool with optimal input data. The angle of the ultrasound probe should also be adjusted so that the entire LV is visible and the image angle remains stable throughout the perioperative phase. The TEE approach is more advantageous than TTE in maintaining a steady angle. The setup time of the tool is short because it utilizes the imaging modalities that are already in use during current cardiac surgery procedures. An automated cardiac monitoring system should not cause harm to the patient, and the tool's use of hardware already used during surgery does not harm the patient. Another requirement for such a system that the tool fulfills is a rapid response time.

It is a common problem in clinical ML that the dataset can have a skewed distribution of health conditions, thus introducing biases to the model. The interviews revealed that this is not an issue with the medical dataset provided by St. Olavs Hospital due to the wide variety of heart conditions in the patient population of the data. Another common issue concerning the data is noise, which can reduce the accuracy of a ML model if the training data is noisy. However, if the TEE video frames are obscured, the tracking target is not annotated, and the model is not trained on that specific frame. Additionally, TransformerTrack reduces the risk of tracking the wrong points in a frame by paying attention to the temporal information among successive frames. By exploiting the temporal information, the frames are mutually reinforced.

The correctness of the tool's output depends both on the ML 's accuracy of the cardiac function estimates and the value of the threshold. In the event of a FP outcome an echocardiographer expert must unnecessarily be called. Consequently, additional time is spent per surgery, which affects the number of patients the hospital can treat each day. In an economic sense, it is expensive for the hospital to demand an unnecessary workforce and increase the daily number of patients undergoing surgery. On the other hand, a high FN ratio may lead to mortality in the worst case, which will be highly costly for the hospital. If several patients lose years of assumed good health to morbidity or mortality, emphasizing FP outcomes might be more financially wise. The FP or FN ratio must be based on how the hospital has allocated its resources and the patient safety. There must be a balance in the outputs concerning the patient's potential health loss and the expenses of false alarms with unnecessary delays.

5.1.2   RQ2: How will the adapted visual tracking model, TransformerTrack, perform on TEE images?

The network with the best value of the mean intersection over union is Network 3. Network 3 also performed best on the mean distance measure. Network 5 showed relatively poor performance, especially on the mean distance measure. When interpreting the measures of Network 5 together, it appears like it rarely manages to track well, but it sometimes hits with high overlap. The performance of Network 5 seems random, in comparison to the other networks. This shows that the TransformerTrack model learns from the ultrasound images in the training

datasets provided by St. Olavs Hospital.

The TransformerTrack shows promising results in the tracking of medical ultrasound images. The best performing model was trained on a combination of manually-and automatically annotated data, with 2D TEE video frames transformed from 3D video frames, with a bounding box size of 15 pixels, and tracked the mitral points. The network produced tracking with a mean distance between the predicted point and the reference points of 5.565 +− 4.763 pixels. Although the tracking performance measures seem promising, the measures alone do not say anything for sure. The measures should be seen relative to other methods' results or the physicians' manual performance. The tracking performance of previously unseen points was poorer than the tracking performance of the same points as in the training data, based on both the qualitative and quantitative performance measures. The lower performance on tracking myocardial segments is as expected, since the model is not trained on such cases.

The training performance with the large bounding box of 25 pixels proved better than the bounding box of 5 and 15 pixels, and the small bounding box performed the poorest. However, the validation loss does decrease with a bounding box of size 5. The table 4.2 shows that the network with the bounding box size of 15 pixels produces better tracking results than with 5 or 25 pixels.

The validation loss did not show any significant changes in the first epochs. The validation loss is high and relatively constant in figure 4.5, 4.6, 4.7, and 4.8. The high validation loss indicates that the models perform poorly on previously unseen data, meaning that the TransformerTrack model is overfitting. From the manual assessment of the loss plots, the model appear to overfit to the same extent regardless of the training dataset.

For each training epoch, one can save a version of the model with corresponding weights. Figure 2.1 shows that the ideal range of model complexity can range over several epochs, making it counterintuitive to choose the optimal model regarding overfitting and underfitting. The best model is most likely inside that interval; however, it varies for each training problem, and there is no correct answer. The plot in the figure 2.16 shows that the first epoch of the ideal range starts where the validation loss is at its lowest, which can be easy to find. However, the model has seen more training data in the descendant epochs, and the training loss will be reduced. There will then be a trade-off between how much lower the training loss is versus how much the validation loss has increased since its minimum. All the validation plots in the experiments were high and relatively constant, thus making it hard to find the minimum of the loss and the best network weights. Therefore, it could be wise to test the tracking with the network on several epochs, for instance, epochs 5, 25, and 50.

5.1.3 Analysis of the Work

The tracking results seem correct and reproducible, based on the fact that the performance measures on the different networks do not differ from one another to a large extent. The automatic annotation is manually corrected by the author. However, the author does not have

profound knowledge of cardiac anatomy. The annotation of the myocardial points done by the author might be inaccurate due to the lack of domain knowledge.

The interviews provided adequate information about the perioperative phase of cardiac surgery in Trondheim regarding work processes, time use, severity, and potential risk.

The findings were not in conflict with anything found in the literature. Commonly, complex models can overfit when trained on relatively small datasets.

The objectives presented in section 1.2 are broken down to more nuanced sub-objectives, and the implementation is commented on:

- Investigate which physicians the ML-based tool will affect the most:
  this was successfully identified during the interviews
- Get an overview of the processes in the OR during the perioperative phase:
  this was successfully identified during the interviews
- Identify which phases of the surgery are most critical for the patient:
  this was successfully identified during the interviews
- Identify which operation phases where the tool can provide the most benefit:
  this was successfully identified during the interviews
- Analyse how the tool will affect these phases of the surgery:
  this was done in the creation of the To-Be process mapping
- Develop an As-Is and To-Be process map:
  this was done based on the interviews
- Identify the advantages and challenges of using ML in a cardiac clinical domain, and if the deployment of the tool will face the challenges:
  this was done based on the literature review and the interviews
- Map the possible outcomes of the model:
  this was done based on the literature review and the interviews and presented in subsection 4.1
- Manually validate the annotation:
  successfully done
- Get an overview of relevant servers and libraries:
  successfully done, the findings are to be found in the appendix B and appendix A
- Calculate the performance of the training of the TransformerTrack model on the medical datasets:
  successfully done, and presented in section 4
- Create a program to represent the tracking performance quantitatively:
  successfully done, the results are presented in section 4
  - Measure the mean distance from the reference landmarks with the models' landmark prediction
  - Compute the standard deviation of the distance
  - Measure the mean intersection over union
  - Measure the percentage of landmarks that are not tracked

- Create a program to represent the tracking performance with images and video: **successfully done**
- Create a data loading program so that the medical data transforms into the right format: **successfully done**
- Explore preprocessing needed for the medical data to improve the model's performance: **the exploration is done based on the relevant literature, but the implementation and testing are not yet carried out. Section 5.2 will present the findings.**
- Create a landmark annotation program for the raw ultrasound data **successfully done and tested**

5.2   Limitations and Further Work

Due to time limitations, not all parameter tuning was tested during training for Transformer-Track. Additionally, the strain value based on the tracked points was not computed. Further work should be done on measuring the longitudinal strain value, based on the tracking results. For the computation of strain, with equation 2.1, the predicted landmarks will be used. The predicted location of the landmarks must be extracted from the predicted bounding box position, as illustrated in figure 3.4. Based on relevant literature on similar tracking problems, some actions and hyperparameter tuning are suggested to improve the model and mitigate overfitting:

- Change the optimizer
- Test networks on several epochs
- Use data augmentation techniques for more diversified data
- Increase the number of epochs, as long as the loss keeps decreasing
- Change the learning rate
- Suppress noise (filter the frame by, for instance, a 7x7 median filter)
- Highlight edges
- Change pixel brightness
- Enhance contrasts
- Emphasize bright neighborhoods

The selection of interviewees is a limitation of this work. Doctors from other workplaces may experience an OR that is organized differently. Including doctors from other hospitals as interviewees may contribute to more general and thorough analysis. If the research is continued, doctors from other hospitals should be included in the interviews.

It can be challenging for outsiders of the APOC project to gain insight into the software supply chain since the software is built on components from different libraries and packages. Consequently, it is hard to get an overview of the potential risks or software license compliance. A software bill of materials (SBOM) can be made to provide visibility to these areas.

When developing a decision-support tool with a loss function, one could consider the FP and FN ratio when calibrating the loss function. However, this was not relevant for the technical work in this project. In the future, the research group of APOC can consider whether the devel-

opers should tune the algorithm to reduce the FP rate or the FN rate. In the assessment, they should take several aspects into account, including the economy of the workforce, the cost of patient mortality, and the potential patient risks. Additionally, the variability of the cardiac function to be estimated should be considered. The displacement of the mitral valvular plane is quite general, but the longitudinal strain is a patient-dependent value. When assessing these factors, the understanding of the tool's role in the OR is crucial.

The developers can make an upper and lower boundary for the threshold value to reduce the chances for human errors when setting the threshold for the tool. Domain experts should, in that case, determine the boundaries of the threshold value.

After the automated tool is developed and ready to use in the OR, the parameters on the ultrasound machine must be tested to find the settings that provide maximum performance. Moreover, the echocardiographer experts must work accordingly with the monitoring tool in use during future surgeries.

The technology can be designed to reduce the over-reliance on the tool. The tool's output will be independent of the threshold, and in the future, the output curve may also include the confidence of the measures. Then the doctors can take the confidence into account when the alarm is raised and possibly reduce the FP ratio. Additionally, it may reduce the FN ratio if the physicians actively assess the tool's output curve.

Further work on tracking ultrasound data with TransformerTrack can include validating the performance based on comparison with other models. Then, different methods should track points in the same test set, and the performance measures in section 4.2.4 should be calculated. Another way to put the tracking results in context is to get doctors to place a bounding box on the same images and compare the performance.

The developers can implement functionality that facilitates continuous learning to improve the model as it functions. Then the doctors need to validate if the model performs correctly, and the technicians must train the model on the new data.

It will be wise to meassure TP, TN, FP and FN of noise predictions, and use it for accuracy- and precision measures.

Chapter 6

Conclusion

To conclude, this project contributed to the ML-based automation of perioperative monitoring of the cardiac function, both at the technical level and the process level. Both RQ1 and RQ2 were answered in section 5, and all the objectives were met. Limitations of the work and suggestions to further work were presented in section 5.2.

A supervised DL model was used for visual object tracking of TEE images. The network architecture named TransformerTrack was used to utilize the temporal information among the frames in the ultrasound video sequences for tracking. The model was trained on 2D TEE video sequence frames from patients at St. Olavs Hospital with landmark annotation of the mitral points. TransformerTrack showed promising results on previously seen data. However, the model is poor at generalizing. The model trained on the auto-annotated dataset, Dataset 3, with 50 epochs and a bounding box of 15 pixels performed best on the tracking of previously seen mitral points. The mean intersection over union was 0.432, and the mean distance between the models prediction and the ground truth was 5.565 +− 4.763 pixels. Although these results seem promising, the tracking results should, in further work, be compared to other methods.

The As-Is and To-Be process mappings, presented in section 4.1, were based on interviews with doctors and a literature review. The mappings illustrate how the monitoring tool will affect the work processes in the OR. The tool will affect the time spent, the communication, patient safety, and the workflow in the pre-and post-operative phase. The changes in safety and surgery time depend on the FP and FN ratio. The model's accuracy, the threshold value, and the hospital's allocation of resources will affect the ratio.

The tool allows for utilization of TEE images for monitoring purposes and automatic estimation of quantitative cardiac function measures. Consequently, the monitoring tool will reduce intra- and intervariability. Additionally, the workload for the physicians will decrease, giving time for other tasks. The physician's performance may be reduced if they heavily rely on the tool and the output is FP or FN. Suggestions were made to implement changes in the tool's functionality to decrease the risks of over-reliance. Furthermore, suggestions were made to ensure an acceptable FP or FN ratio.

Bibliography

[1]   W. H. Organization, *Global status report on noncommunicable diseases 2010*, 2011.

[2]   P. Ponikowski, A. A. Voors, S. D. Anker, H. Bueno, J. G. F. Cleland, A. J. S. Coats, V. Falk, J. R. González-Juanatey, V.-P. Harjola, E. A. Jankowska, and et al., "2016 esc guidelines for the diagnosis and treatment of acute and chronic heart failure," *European Journal of Heart Failure*, vol. 18, no. 8, pp. 891–975, May 2016. DOI: `10.1002/ejhf.592`. [Online]. Available: `https://onlinelibrary.wiley.com/doi/10.1002/ejhf.592`.

[3]   E. T. Aasheim, "Hvor stor er risikoen ved hjerteoperasjon?" nor, *Tidsskrift for den Norske Lægeforening,* vol. 131, no. 17, pp. 1629–1629, 2011, ISSN: 0029-2001.

[4]   2019. [Online]. Available: `https://helse-midt.no/nyheter/2020/les-mer-om-prosjektene-som-fikk-innovasjonsmidler-i-2019`.

[5]   "Intraoperative interventions: American college of chest physicians guidelines for the prevention and management of postoperative atrial fibrillation after cardiac surgery.," *Proquest.com*, 2019. DOI: `",`. [Online]. Available: `https://www.proquest.com/docview/200452102`.

[6]   J.-L. Vincent, P. Pelosi, R. Pearse, D. Payen, A. Perel, A. Hoeft, S. Romagnoli, V. M. Ranieri, C. Ichai, P. Forget, G. D. Rocca, and A. Rhodes, "Perioperative cardiovascular monitoring of high-risk patients: A consensus of 12," *Critical Care*, vol. 19, no. 1, Dec. 2015. DOI: `10.1186/s13054-015-0932-7`. [Online]. Available: `https://ccforum.biomedcentral.com/articles/10.1186/s13054-015-0932-7`.

[7]   A. Rafiq, E. Sklyar, and J. N. Bella, "Cardiac evaluation and monitoring of patients undergoing noncardiac surgery," *Health Services Insights*, vol. 10, Jan. 2017. DOI: `10.1177/1178632916686074`. [Online]. Available: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5398290/`.

[8]   S. T. Reeves, A. C. Finley, N. J. Skubas, M. Swaminathan, W. S. Whitley, K. E. Glas, R. T. Hahn, J. S. Shanewise, M. S. Adams, and S. K. Shernan, "Basic perioperative transesophageal echocardiography examination: A consensus statement of the american society of echocardiography and the society of cardiovascular anesthesiologists," *Journal of the American Society of Echocardiography*, vol. 26, no. 5, pp. 443–456, May 2013. DOI: `10.1016/j.echo.2013.02.015`.

[9]   A. J. Labovitz, V. E. Noble, M. Bierig, S. A. Goldstein, R. Jones, S. Kort, T. R. Porter, K. T. Spencer, V. S. Tayal, and K. Wei, "Focused cardiac ultrasound in the emergent setting: A consensus statement of the american society of echocardiography and american college of emergency physicians," *Journal of the American Society of Echocardiography*, vol. 23, no. 12, pp. 1225–1230, Dec. 2010. DOI: `10.1016/j.echo.2010.10.005`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0894731710008710?casa_token=8l8ASD3g3h4AAAAA:t7CqHtt7cGRsUa-SYOgwLtRrRD_X8buq6i9q9kiPeCCpacIm1uhqA-dCcXN_rq9kGanbYCM5`.

[10] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, and T. Wang, "Deep learning in medical ultrasound analysis: A review," *Engineering*, vol. 5, no. 2, pp. 261–275, 2019, ISSN: 2095-8099. DOI: `https://doi.org/10.1016/j.eng.2018.11.020`.

[11] A. Poddar, H. S. Yang, C. Padmanabhan, J. Maalouf, and K. Chandrasekaran, "Intra-operative trans-esophageal echocardiography in heart valve disease," *Indian Journal of Thoracic and Cardiovascular Surgery*, vol. 36, no. S1, pp. 140–153, Jan. 2020. DOI: `10.1007/s12055-019-00909-9`. [Online]. Available: `https://pubmed.ncbi.nlm.nih.gov/33061195/`.

[12] J.-L. Vincent, A. Rhodes, A. Perel, G. S. Martin, G. Rocca, B. Vallet, M. R. Pinsky, C. K. Hofer, J.-L. Teboul, W.-P. de Boode, S. Scolletta, A. Vieillard-Baron, D. De Backer, K. R. Walley, M. Maggiorini, and M. Singer, "Clinical review: Update on hemodynamic monitoring - a consensus of 16," *Critical Care*, vol. 15, no. 4, p. 229, 2011. DOI: `10.1186/cc10291`. [Online]. Available: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3387592/`.

[13] D. G. Lappas, J. John Powell W. M., and W. M. Daggett, "Cardiac Dysfunction in the Perioperative Period: Pathophysiology, Diagnosis, and Treatment," *Anesthesiology*, vol. 47, no. 2, pp. 117–137, Aug. 1977, ISSN: 0003-3022. DOI: `10.1097/00000542-197708000-00004`. eprint: `https://pubs.asahq.org/anesthesiology/article-pdf/47/2/117/297752/0000542-197708000-00004.pdf`. [Online]. Available: `https://doi.org/10.1097/00000542-197708000-00004`.

[14] H. Dalen, A. Thorstensen, S. A. Aase, C. B. Ingul, H. Torp, L. J. Vatten, and A. Stoylen, "Segmental and global longitudinal strain and strain rate based on echocardiography of 1266 healthy individuals: The hunt study in norway," *European journal of echocardiography*, vol. 11, no. 2, pp. 176–183, 2010.

[15] O. A. Smiseth, H. Torp, A. Opdahl, K. H. Haugaa, and S. Urheim, "Myocardial strain imaging: How useful is it in clinical decision making?" *European Heart Journal*, vol. 37, no. 15, pp. 1196–1207, Oct. 2015. DOI: `10.1093/eurheartj/ehv529`. [Online]. Available: `https://pubmed.ncbi.nlm.nih.gov/26508168/`.

[16] A. Scatteia, A. Baritussio, and C. Bucciarelli-Ducci, "Strain imaging using cardiac magnetic resonance," *Heart Failure Reviews*, vol. 22, no. 4, pp. 465–476, Jun. 2017. DOI: `10.1007/s10741-017-9621-8`. [Online]. Available: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5487809/`.

[17] M. Tee, J. A. Noble, and D. A. Bluemke, "Imaging techniques for cardiac strain and deformation: Comparison of echocardiography, cardiac magnetic resonance and cardiac computed tomography," *Expert Review of Cardiovascular Therapy*, vol. 11, no. 2, pp. 221–231, 2013, PMID: 23405842. DOI: `10.1586/erc.12.182`. [Online]. Available: `https://doi.org/10.1586/erc.12.182`.

[18] E. Altiok, S. Tiemann, M. Becker, R. Koos, C. Zwicker, J. Schroeder, N. Kraemer, F. Schoth, D. Adam, Z. Friedman, N. Marx, and R. Hoffmann, "Myocardial deformation imaging by two-dimensional speckle-tracking echocardiography for prediction of global and segmental functional changes after acute myocardial infarction: A comparison with late gadolinium enhancement cardiac magnetic resonance," *Journal of the American Society of Echocardiography*, vol. 27, no. 3, pp. 249–257, Mar. 2014. DOI: `10.1016/j.echo.2013.11.014`. [Online]. Available: `https://pubmed.ncbi.nlm.nih.gov/24368027/`.

[19] S. J. Buss, F. Schulz, D. Mereles, W. Hosch, C. Galuschky, G. Schummers, D. Stapf, N. Hofmann, E. Giannitsis, S. E. Hardt, H.-U. Kauczor, H. A. Katus, and G. Korosoglou, "Quantitative analysis of left ventricular strain using cardiac computed tomography," *European Journal of Radiology*, vol. 83, no. 3, e123–e130, Mar. 2014. DOI: `10.1016/j.ejrad.2013.11.026`. [Online]. Available: `https://pubmed.ncbi.nlm.nih.gov/24368011/`.

[20] P. Collier, D. Phelan, and A. Klein, "A test in context: Myocardial strain measured by speckle-tracking echocardiography," *Journal of the American College of Cardiology*, vol. 69, no. 8, pp. 1043–1056, Feb. 2017. DOI: `10.1016/j.jacc.2016.12.012`.

[21] J. D'hooge, E. Konofagou, F. Jamal, A. Heimdal, L. Barrios, B. Bijnens, J. Thoen, F. Van de Werf, G. Sutherland, and P. Suetens, "Two-dimensional ultrasonic strain rate measurement of the human heart in vivo," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 49, no. 2, pp. 281–286, Feb. 2002. DOI: `10.1109/58.985712`. [Online]. Available: `https://ieeexplore.ieee.org/abstract/document/985712`.

[22] J. D'hooge, "Regional strain and strain rate measurements by cardiac ultrasound: Principles, implementation and limitations," *European Journal of Echocardiography*, vol. 1, no. 3, pp. 154–170, Sep. 2000. DOI: `10.1053/euje.2000.0031`. [Online]. Available: `https://pubmed.ncbi.nlm.nih.gov/11916589/`.

[23] A. Støylen, A. Heimdal, K. Bjørnstad, R. Wiseth, H. Vik-Mo, H. Torp, B. Angelsen, and T. Skjærpe, "Strain rate imaging by ultrasonography in the diagnosis of coronary artery disease," *Journal of the American Society of Echocardiography*, vol. 13, no. 12, pp. 1053–1064, 2000.

[24] S. Urheim, T. Edvardsen, H. Torp, B. Angelsen, and O. A. Smiseth, "Myocardial strain by doppler echocardiography: Validation of a new method to quantify regional myocardial function," *Circulation*, vol. 102, no. 10, pp. 1158–1164, 2000.

[25] T. H. Marwick, "Measurement of strain and strain rate by echocardiography," *Journal of the American College of Cardiology*, vol. 47, no. 7, pp. 1313–1327, Apr. 2006. DOI: `10.1016/j.jacc.2005.11.063`.

[26] B. H. Amundsen, T. Helle-Valle, T. Edvardsen, H. Torp, J. Crosby, E. Lyseggen, A. Støylen, H. Ihlen, J. A. Lima, O. A. Smiseth, and S. A. Slørdahl, "Noninvasive myocardial strain measurement by speckle tracking echocardiography," *Journal of the American College of Cardiology*, vol. 47, no. 4, pp. 789–793, 2006. DOI: `10.1016/j.jacc.2005.10.040`. [Online]. Available: `https://www.jacc.org/doi/abs/10.1016/j.jacc.2005.10.040`.

[27] A. Østvik, E. Smistad, T. Espeland, E. A. R. Berg, and L. Lovstakken, "Automatic myocardial strain imaging in echocardiography using deep learning," *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 309–316, 2018. DOI: `10.1007/978-3-030-00889-5_35`. [Online]. Available: `https://link.springer.com/chapter/10.1007/978-3-030-00889-5_35`.

[28] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational Intelligence and Neuroscience*, vol. 2018, pp. 1–13, 2018. DOI: `10.1155/2018/7068349`. [Online]. Available: `https://www.hindawi.com/journals/cin/2018/7068349/`.

[29] R. Szeliski, *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[30] F. A. Flachskampf, T. Biering-Sørensen, S. D. Solomon, O. Duvernoy, T. Bjerner, and O. A. Smiseth, "Cardiac imaging to evaluate left ventricular diastolic function," *JACC: Cardiovascular Imaging*, vol. 8, no. 9, pp. 1071–1093, 2015.

[31] J. Xing, H. Ai, and S. Lao, *Multiple human tracking based on multi-view upper-body detection and discriminative learning*, Aug. 2010. [Online]. Available: `https://www.researchgate.net/publication/220932068_Multiple_Human_Tracking_Based_on_Multi-view_Upper-Body_Detection_and_Discriminative_Learning`.

[32] G. Zhang and P. A. Vela, "Good features to track for visual slam," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1373–1382.

[33] L. Liu, J. Xing, H. Ai, and X. Ruan, "Hand posture recognition using finger geometric feature," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, IEEE, 2012, pp. 565–568.

[34] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "Siamrpn++: Evolution of siamese visual tracking with very deep networks," *arXiv.org*, 2018. DOI: `10.48550/arXiv.1812.11703`. [Online]. Available: `https://arxiv.org/abs/1812.11703`.

[35] C. Paton and S. Kobayashi, "An open science approach to artificial intelligence in healthcare," *Yearbook of Medical Informatics*, vol. 28, no. 01, pp. 047–051, Apr. 2019. DOI: `10.1055/s-0039-1677898`. [Online]. Available: `https://www.thieme-connect.com/products/ejournals/html/10.1055/s-0039-1677898`.

[36] C. E. Bouton, A. Shaikhouni, N. V. Annetta, M. A. Bockbrader, D. A. Friedenberg, D. M. Nielson, G. Sharma, P. B. Sederberg, B. C. Glenn, W. J. Mysiw, *et al.*, "Restoring cortical control of functional movement in a human with quadriplegia," *Nature*, vol. 533, no. 7602, pp. 247–250, 2016.

[37] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[38] M. Alkhatib, A. Hafiane, P. Vieyres, and A. Delbos, "Deep visual nerve tracking in ultrasound images," *Computerized Medical Imaging and Graphics*, vol. 76, p. 101 639, 2019.

[39] W. G. Daniel and A. Mügge, "Transesophageal echocardiography," *New England Journal of Medicine*, vol. 332, no. 19, pp. 1268–1280, May 1995. DOI: `10.1056/nejm199505113321906`.

[40] J. B. SEWARD, B. K. KHANDHERIA, J. K. OH, M. D. ABEL, R. W. HUGHES, W. D. ED-WARDS, B. A. NICHOLS, W. K. FREEMAN, and A. J. TAJIK, "Transesophageal echocardiography: Technique, anatomic correlations, implementation, and clinical applications," *Mayo Clinic Proceedings*, vol. 63, no. 7, pp. 649–680, Jul. 1988. DOI: `10.1016/s0025-6196(12)65529-3`. [Online]. Available: `https://www.sciencedirect.com/science/article/abs/pii/S0025619612655293`.

[41] E. J. Gussenhoven, M. A. Taams, J. R. Roelandt, K. M. Ligtvoet, J. McGhie, L. A. van Herwerden, and M. K. Cahalan, "Transesophageal two-dimensional echocardiography: Its role in solving clinical problems," *Journal of the American College of Cardiology*, vol. 8, no. 4, pp. 975–979, Oct. 1986. DOI: `10.1016/s0735-1097(86)80444-2`. [Online]. Available: `https://pubmed.ncbi.nlm.nih.gov/3760371/`.

[42] K. Hu, D. Liu, S. Herrmann, M. Niemann, P. D. Gaudron, W. Voelker, G. Ertl, B. Bijnens, and F. Weidemann, "Clinical implication of mitral annular plane systolic excursion for patients with cardiovascular disease," *European Heart Journal–Cardiovascular Imaging*, vol. 14, no. 3, pp. 205–212, 2013.

[43] M. Sengeløv, P. G. Jørgensen, J. S. Jensen, N. E. Bruun, F. J. Olsen, T. Fritz-Hansen, K. Nochioka, and T. Biering-Sørensen, "Global longitudinal strain is a superior predictor of all-cause mortality in heart failure with reduced ejection fraction," *JACC: Cardiovascular Imaging*, vol. 8, no. 12, pp. 1351–1359, 2015, ISSN: 1936-878X. DOI: `https://doi.org/10.1016/j.jcmg.2015.07.013`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S1936878X15007196`.

[44] S. Tonekaboni, S. Joshi, M. D. McCradden, and A. Goldenberg, "What clinicians want: Contextualizing explainable machine learning for clinical end use," in *Machine learning for healthcare conference*, PMLR, 2019, pp. 359–380.

[45] A. D. Bedoya, M. E. Clement, M. Phelan, R. C. Steorts, C. O'Brien, and B. A. Goldstein, "Minimal impact of implemented early warning score and best practice alert for patient deterioration," *Critical care medicine*, vol. 47, no. 1, p. 49, 2019.

[46] J. L. Guidi, K. Clark, M. T. Upton, H. Faust, C. A. Umscheid, M. B. Lane-Fall, M. E. Mikkelsen, W. D. Schweickert, C. A. Vanzandbergen, J. Betesh, *et al.*, "Clinician perception of the effectiveness of an automated early warning and response system for sepsis in an academic medical center," *Annals of the American Thoracic Society*, vol. 12, no. 10, pp. 1514–1519, 2015.

[47]   2019. [Online]. Available: https://www.sintef.no/ekspertise/sintef-ikt/teknologiledelse/produksjonsledelse/.

[48]   S. Franken, S. Kolvenbach, W. Prinz, I. Alvertis, and S. Koussouris, "Cloudteams: Bridging the gap between developers and customers during software development processes," *Procedia Computer Science*, vol. 68, pp. 188–195, 2015, 1st International Conference on Cloud Forward: From Distributed to Complete Computing, ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2015.09.234.

[49]   N. Wang, W. Zhou, J. Wang, and H. Li, "Transformer meets tracker: Exploiting temporal context for robust visual tracking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1571–1580.

[50]   M. Elsherbiny, Y. Abdelwahab, K. Nagy, A. Mannaa, and Y. Hassabelnaby, "Role of intraoperative transesophageal echocardiography in cardiac surgery: An observational study," *Open Access Macedonian Journal of Medical Sciences*, vol. 7, no. 15, pp. 2480–2483, Aug. 2019. DOI: 10.3889/oamjms.2019.712. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6814460/.

[51]   J. F. Grue, S. Storve, H. Dalen, Ø. Salvesen, O. C. Mjølstad, S. O. Samstad, H. Torp, and B. O. Haugen, "Automatic measurements of mitral annular plane systolic excursion and velocities to detect left ventricular dysfunction," *Ultrasound in Medicine Biology*, vol. 44, no. 1, pp. 168–176, Jan. 2018. DOI: 10.1016/j.ultrasmedbio.2017.09.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S030156291732272X?via%5C%3Dihub.

[52]   N. senter, *Bedre bruk av kunstig intelligens i norsk spesialisthelsetjeneste*, 2019. [Online]. Available: https://ehealthresearch.no/prosjekter/bedre-bruk-av-kunstig-intelligens-i-norsk-spesialisthelsetjeneste.

[53]   O. Sand, Ø. V. Sjaastad, E. Haug, and J. G. Bjålie, *Menneskekroppen fysiologi og anatomi*. Gyldendal akademisk, 2018, ISBN: 9788205504820.

[54]   F. Paulsen and D. J. Waschke Professor, *Sobotta Atlas of Human Anatomy, Internal Organs, 15th ed., English/Latin*. Churchill Livingstone, 2011, ISBN: 9780723437314.

[55]   L. M. Biga, S. Dawson, A. Harwell, R. Hopkins, J. Kaufmann, M. LeMaster, P. Matern, K. Morrison-Graham, D. Quick, and J. Runyeon, *19.3 cardiac cycle*, Sep. 2019. [Online]. Available: https://open.oregonstate.education/aandp/chapter/19-3-cardiac-cycle/.

[56]   M. Sarkar and V. Prabhu, "Basics of cardiopulmonary bypass," *Indian Journal of Anaesthesia*, vol. 61, no. 9, p. 760, 2017. DOI: 10.4103/ija.ija_379_17. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5613602/.

[57]   L. Bohs and G. Trahey, "A novel method for angle independent ultrasonic imaging of blood flow and tissue motion," *IEEE Transactions on Biomedical Engineering*, vol. 38, no. 3, pp. 280–286, 1991. DOI: 10.1109/10.133210.

[58] D. L. Mann, L. D. Gillam, and A. E. Weyman, "Cross-sectional echocardiographic assessment of regional left ventricular performance and myocardial perfusion," *Progress in Cardiovascular Diseases*, vol. 29, no. 1, pp. 1–52, Jul. 1986. DOI: 10.1016/0033-0620(86)90017-4. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/0033062086900174.

[59] M. Dandel and R. Hetzer, "Echocardiographic strain and strain rate imaging—clinical applications," *International journal of cardiology*, vol. 132, no. 1, pp. 11–24, 2009.

[60] R. Gonzalez, R. Woods, P. Pearson, and Hall, *Digital Image Processing Third Edition Pearson International Edition prepared by Pearson Education*. Pearson, 2010.

[61] C. M. Otto, *Textbook of clinical echocardiography*. Elsevier Health Sciences, 2013.

[62] G. Xiao, M. Brady, J. Noble, and Y. Zhang, "Segmentation of ultrasound b-mode images with intensity inhomogeneity correction," *IEEE Transactions on Medical Imaging*, vol. 21, no. 1, pp. 48–57, 2002. DOI: 10.1109/42.981233.

[63] S. A. Matulevicius, A. Rohatgi, S. R. Das, A. L. Price, A. deLuna, and S. C. Reimold, "Appropriate Use and Clinical Impact of Transthoracic Echocardiography," *JAMA Internal Medicine*, vol. 173, no. 17, pp. 1600–1607, Sep. 2013.

[64] E. M. S. F. BW, *Intraoperative echocardiography*, Sep. 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29763138/.

[65] admin, *2.3.2 apical window*, Jun. 2012. [Online]. Available: https://123sonography.com/ebook/apical-window.

[66] B. D. M. Services, *Bangkok heart hospital*, 2021. [Online]. Available: https://www.bangkokhearthospital.com/en/medical-service/echocardiogram.

[67] P. Blahút, *A4c la pulmonary veins focus (echo)*, 2018. [Online]. Available: https://www.techmed.sk/en/echo/view/24/.

[68] E. Alfnes, *Styringsmodeller, hva er en styringsmodell og hva kan den brukes til?* 2005. [Online]. Available: https://www.ntnu.no/documents/1263511339/1264965608/styringsmodeller.pdf/be693a03-87f2-4c19-80cf-a1b9311da716.

[69] *Industrial Management Data Systems*, vol. 104, no. 8, pp. 637–643, 2022. DOI: 10.1108\/imds.

[70] A. B. Askelund Klepp, M. J. Haugen, and J. O. Strandhagen, *Styringsmodell, Kompendium TPK 4100 Produksjons- og driftsteknikk*. 2010. [Online]. Available: https://dvikan.no/ntnu-studentserver/kompendier/GetFileKompendium.pdf.

[71] J. Bloomberg, "Digitization, digitalization, and digital transformation: Confuse them at your peril," *Forbes. Retrieved on August*, vol. 28, p. 2019, 2018.

[72] P. Solding and P. Gullander, "Concepts for simulation based value stream mapping," in *Proceedings of the 2009 Winter Simulation Conference (WSC)*, 2009, pp. 2231–2237. DOI: 10.1109/WSC.2009.5429185.

[73] J. A. Marin-Garcia, P. I. Vidal-Carreras, and J. J. Garcia-Sabater, "The role of value stream mapping in healthcare services: A scoping review," *International Journal of Environmental Research and Public Health*, vol. 18, no. 3, p. 951, Jan. 2021. DOI: `10.3390/ijerph18030951`. [Online]. Available: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7908358/`.

[74] E. Alfnes, *Styringsmodeller, hva er en styringsmodell og hva kan den brukes til?* 2005. [Online]. Available: `https://www.ntnu.no/documents/1263511339/1264965608/styringsmodeller.pdf/be693a03-87f2-4c19-80cf-a1b9311da716`.

[75] H. Mosley and A. Mayer, "Benchmarking national labour market performance: A radar chart approach," 1999.

[76] C. Matt, T. Hess, and A. Benlian, "Digital transformation strategies," *Business Information Systems Engineering*, vol. 57, no. 5, pp. 339–343, Aug. 2015. DOI: `10.1007/s12599-015-0401-5`. [Online]. Available: `https://link.springer.com/article/10.1007/s12599-015-0401-5#citeas`.

[77] P. Parviainen, M. Tihinen, J. Kääriäinen, and S. Teppola, "Tackling the digitalization challenge: How to benefit from digitalization in practice," *International journal of information systems and project management*, vol. 5, no. 1, pp. 63–77, 2017.

[78] A. Moeuf, R. Pellerin, S. Lamouri, S. Tamayo-Giraldo, and R. Barbaray, "The industrial management of smes in the era of industry 4.0," *International Journal of Production Research*, vol. 56, no. 3, pp. 1118–1136, 2018. DOI: `10.1080/00207543.2017.1372647`. eprint: `https://doi.org/10.1080/00207543.2017.1372647`. [Online]. Available: `https://doi.org/10.1080/00207543.2017.1372647`.

[79] O. E. Oluyisola, F. Sgarbossa, and J. O. Strandhagen, "Smart production planning and control: Concept, use-cases and sustainability implications," *Sustainability*, vol. 12, no. 9, 2020, ISSN: 2071-1050. [Online]. Available: `https://www.mdpi.com/2071-1050/12/9/3791`.

[80] S.-V. Buer, M. Semini, J. O. Strandhagen, and F. Sgarbossa, "The complementary effect of lean manufacturing and digitalisation on operational performance," *International Journal of Production Research*, vol. 59, pp. 1–17, Aug. 2020. DOI: `10.1080/00207543.2020.1790684`.

[81] D. Romero, J. Stahre, T. Wuest, O. Noran, P. Bernus, Å. Fasth Fast-Berglund, and D. Gorecky, "Towards an operator 4.0 typology: A human-centric perspective on the fourth industrial revolution technologies," Oct. 2016.

[82] D. Ivanov, A. Dolgui, A. Das, and B. Sokolov, "Digital supply chain twins: Managing the ripple effect, resilience, and disruption risks by data-driven optimization, simulation, and visibility," *Handbook of Ripple Effects in the Supply Chain*, pp. 309–332, 2019. DOI: `10.1007/978-3-030-14302-2_15`. [Online]. Available: `https://link.springer.com/chapter/10.1007/978-3-030-14302-2_15`.

[83] M. A. Hossain and A. Nadeem, "Towards digitizing the construction industry: State of the art of construction 4.0," in *Proceedings of the ISEC*, vol. 10, 2019.

[84] N. Z. Noor Hasnan and Y. Yusoff, "Short review: Application areas of industry 4.0 technologies in food processing sector," Nov. 2018, pp. 1–6. DOI: 10.1109/SCORED. 2018.8711184.

[85] S. K. Mudda, C. B. Giddi, and P. Murthy, "A study on the digitization of supply chains in agriculture-an indian experience," *Journal of Agricultural Informatics*, vol. 8, no. 1, 2017.

[86] P. F. Borowski, "Digitization, digital twins, blockchain, and industry 4.0 as elements of management process in enterprises in the energy sector," *Energies*, vol. 14, no. 7, p. 1885, 2021.

[87] M. Ghobakhloo, "Industry 4.0, digitization, and opportunities for sustainability," *Journal of cleaner production*, vol. 252, p. 119 869, 2020.

[88] M. Hammer, "Digitization perspective: Impact of digital technologies in manufacturing," in *Management Approach for Resource-Productive Operations*, Springer, 2019, pp. 27–68.

[89] J. G. Carbonell, R. S. Michalski, and T. M. Mitchell, "An overview of machine learning," *Machine learning*, pp. 3–23, 1983.

[90] A. Fiane, J. Bjørnstad, and O. Geiran, "Norsk hjertekirurgiregister, årsrapport for 2020 med plan for forbedringstiltak,"

[91] M. Drummond, B. O'Brien, G. Torrance, and G. Stoddart, *Methods for the Economic Evaluation of Health Care Programmes*, Undefined/Unknown, 2nd. United Kingdom: Oxford University Press, 1997.

[92] I. S. Kristiansen and I. for, *Bør medisinske feil unngås for enhver pris?* Oct. 2000. [Online]. Available: https://tidsskriftet.no/2000/10/kronikk/bor-medisinske-feil-unngas-enhver-pris.

[93] Helsedirektoratet, "Prioritering på klinisk nivå," *Helsedirektoratet*, Nov. 2018. [Online]. Available: https://www.helsedirektoratet.no/tema/prioritering-i-helsetjenesten.

[94] jonas, *Prisen på "et godt leveår": Slik prioriterer det norske helsevesenet*, Apr. 2020. [Online]. Available: https://filternyheter.no/prisen-pa-et-liv-slik-prioriterer-det-norske-helsevesenet/.

[95] H. og omsorgsdepartementet, *Åpent og rettferdig – prioriteringer i helsetjenesten*, Nov. 2014. [Online]. Available: https://www.regjeringen.no/no/dokumenter/NOU-2014-12/id2076730/.

[96] G. Singh, S. J. Al'Aref, M. Van Assen, T. S. Kim, A. van Rosendael, K. K. Kolli, A. Dwivedi, G. Maliakal, M. Pandey, J. Wang, and et al., "Machine learning in cardiac ct: Basic concepts and contemporary data," *Journal of Cardiovascular Computed Tomography*, vol. 12, no. 3, pp. 192–201, May 2018. DOI: 10.1016/j.jcct.2018.04.010.

[97] T. M. Mitchell, *Machine learning, International Edition*, ser. McGraw-Hill Series in Computer Science. McGraw-Hill, 1997.

[98]   C. Rudin and K. L. Wagstaff, *Machine Learning*, vol. 95, no. 1, pp. 1–9, Nov. 2013. DOI: `10.1007/s10994-013-5425-9`. [Online]. Available: `https://link.springer.com/article/10.1007/s10994-013-5425-9`.

[99]   S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6. DOI: `10.1109/ICEngTechnol.2017.8308186`.

[100]   A. Jain, J. Mao, and K. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, 1996. DOI: `10.1109/2.485891`.

[101]   S. J. Russell and P. Norvig, *Artificial Intelligence: a modern approach*, 3rd ed. Pearson, 2009.

[102]   I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, `http://www.deeplearningbook.org`.

[103]   J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, May 2016. DOI: `10.1186/s13634-016-0355-x`. [Online]. Available: `https://asp-eurasipjournals.springeropen.com/articles/10.1186/s13634-016-0355-x`.

[104]   *Ai vs. machine learning vs. deep learning vs. neural networks: What's the difference?*,2020, May 2020. [Online]. Available: `https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks`.

[105]   Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. DOI: `10.1109/5.726791`.

[106]   S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6. DOI: `10.1109/ICEngTechnol.2017.8308186`.

[107]   S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6. DOI: `10.1109/ICEngTechnol.2017.8308186`.

[108]   R. Gonzalez, *Digital image processing. eng. 4th ed.* New York: Pearson, 2018, ISBN: 2018. ISBN: 9781292223049.

[109]   2012. [Online]. Available: `https://cs231n.github.io/convolutional-networks/`.

[110]   R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Jun. 2018. DOI: `10.1007/s13244-018-0639-9`. [Online]. Available: `https://insightsimaging.springeropen.com/articles/10.1007/s13244-018-0639-9`.

[111]   2012. [Online]. Available: `https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8308186`.

[112]   P. Ramachandran, B. Zoph, and Q. V. Le, *Searching for activation functions*, 2017. [Online]. Available: `https://arxiv.org/abs/1710.05941`.

[113] X. Xie, D. Du, Q. Li, Y. Liang, W. T. Tang, Z. L. Ong, M. Lu, H. P. Huynh, and R. S. M. Goh, "Exploiting sparsity to accelerate fully connected layers of cnn-based applications on mobile socs," *ACM Trans. Embed. Comput. Syst.*, vol. 17, no. 2, Dec. 2017, ISSN: 1539-9087. DOI: `10.1145/3122788`. [Online]. Available: `https://doi.org/10.1145/3122788`.

[114] S. R. Yoshida, *Computer Vision*. 2011. [Online]. Available: `https://web.s.ebscohost.com/ehost/ebookviewer/ebook/bmxlYmtfXzU0MTM1OF9fQU41?sid=0a70b445-bb5a-4ab1-b996-2da21751614f@redis&vid=0&format=EB&rid=1`.

[115] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016.

[116] I. Mihajlovic, *Everything you ever wanted to know about computer vision*. Apr. 2019. [Online]. Available: `https://towardsdatascience.com/everything-you-ever-wanted-to-know-about-computer-vision-heres-a-look-why-it-s-so-awesome-e8a58dfb641e`.

[117] Y. Wang, X. Wei, X. Tang, J. Wu, and J. Fang, "Response map evaluation for rgbt tracking," *Neural Computing and Applications*, vol. 34, no. 7, pp. 5757–5769, Jan. 2022. DOI: `10.1007/s00521-021-06704-1`.

[118] M. Fiaz, A. Mahmood, S. Javed, and S. K. Jung, "Handcrafted and deep trackers: Recent visual object tracking approaches and trends," *ACM Comput. Surv.*, vol. 52, no. 2, Apr. 2019, ISSN: 0360-0300. DOI: `10.1145/3309665`. [Online]. Available: `https://doi.org/10.1145/3309665`.

[119] J. Kwon and K. M. Lee, "Visual tracking decomposition," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2010. DOI: `10.1109/cvpr.2010.5539821`. [Online]. Available: `https://ieeexplore.ieee.org/abstract/document/5539821`.

[120] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and Philip, *Fully-convolutional siamese networks for object tracking*, 2016. [Online]. Available: `https://arxiv.org/abs/1606.09549`.

[121] G. Bhat, M. Danelljan, L. V. Gool, and R. Timofte, "Learning discriminative model prediction for tracking," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019.

[122] A. Vaswani, G. Brain, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, Ł. Kaiser, and I. Polosukhin, *Attention Is All You Need*. [Online]. Available: `https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf`.

[123] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, 2015. [Online]. Available: `https://arxiv.org/abs/1512.03385`.

[124] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, "A survey on vision transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2022. DOI: `10.1109/TPAMI.2022.3152247`.

[125]	L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and Philip, *Fully-convolutional siamese networks for object tracking*, 2016. [Online]. Available: `https://arxiv.org/abs/1606.09549`.

[126]	G. Bhat, M. Danelljan, L. V. Gool, and R. Timofte, "Learning discriminative model prediction for tracking," *Thecvf.com*, pp. 6182–6191, 2019. [Online]. Available: `https://openaccess.thecvf.com/content_ICCV_2019/html/Bhat_Learning_Discriminative_Model_Prediction_for_Tracking_ICCV_2019_paper.html`.

[127]	T. Shah, *About train, validation and test sets in machine learning*, Dec. 2017. [Online]. Available: `https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7`.

[128]	E. N. Minor, S. D. Howard, A. A. Green, M. A. Glaser, C. S. Park, and N. A. Clark, "End-to-end machine learning for experimental physics: Using simulated data to train a neural network for object detection in video microscopy," *Soft matter*, vol. 16, no. 7, pp. 1751–1759, 2020.

[129]	T. Sasakawa, J. Hu, and K. Hirasawa, "A brainlike learning system with supervised, unsupervised, and reinforcement learning," *Electrical Engineering in Japan*, vol. 162, no. 1, pp. 32–39, 2007. DOI: 10.1002/eej.20600. [Online]. Available: `https://onlinelibrary.wiley.com/doi/abs/10.1002/eej.20600`.

[130]	S. Krig, "Ground truth data, content, metrics, and analysis," in *Computer Vision Metrics*, Springer, 2016, pp. 247–271.

[131]	D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognition*, vol. 45, no. 1, pp. 346–362, Jan. 2012. DOI: 10.1016/j.patcog.2011.05.013. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0031320311002391?casa_token=9eqUU1rvVj8AAAAA:qEm6a9IVddCHCbSx3XTCD_ZjCVa5tlMrD-BGbF1d73MuDnfOpHNe4kArFyLTaQPtvcXE_K3I`.

[132]	T. Hastie, R. Tibshirani, and J. Friedman, "Overview of supervised learning," *The Elements of Statistical Learning*, pp. 9–41, Dec. 2008. DOI: 10.1007/978-0-387-84858-7_2. [Online]. Available: `https://link.springer.com/chapter/10.1007%5C%2F978-0-387-84858-7_2`.

[133]	B. Liu, "Lifelong machine learning: A paradigm for continuous learning," *Frontiers of Computer Science*, vol. 11, no. 3, pp. 359–361, 2017.

[134]	L. Torrey and J. Shavlik, "Transfer learning," *Handbook of Research on Machine Learning Applications and Trends*, pp. 242–264, 2010. DOI: 10.4018/978-1-60566-766-9.ch011. [Online]. Available: `https://www.igi-global.com/chapter/transfer-learning/36988`.

[135]	W. Tao, M. Al-Amin, H. Chen, M. Leu, Z. Yin, and R. Qin, "Real-time assembly operation recognition with fog computing and transfer learning for human-centered intelligent manufacturing," *Procedia Manufacturing*, vol. 48, pp. 926–931, Jan. 2020. DOI: 10.1016/j.promfg.2020.05.131.

[136] M. Danelljan, L. V. Gool, and R. Timofte, "Probabilistic regression for visual tracking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7183–7192.

[137] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, Mar. 1951. DOI: `10.1214/aoms/1177729694`. [Online]. Available: `https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-22/issue-1/On-Information-and-Sufficiency/10.1214/aoms/1177729694.full`.

[138] J. R. Hershey and P. A. Olsen, "Approximating the kullback leibler divergence between gaussian mixture models," *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, Apr. 2007. DOI: `10.1109/icassp.2007.366913`.

[139] A. Culotta, P. Kanani, R. Hall, M. Wick, and A. McCallum, "Author disambiguation using error-driven machine learning with a ranking loss function," in *Sixth International Workshop on Information Integration on the Web (IIWeb-07), Vancouver, Canada*, 2007.

[140] J. Xu, Z. Zhang, T. Friedman, Y. Liang, and G. Van den Broeck, "A semantic loss function for deep learning with symbolic knowledge," in *Proceedings of the 35th International Conference on Machine Learning*, J. Dy and A. Krause, Eds., ser. Proceedings of Machine Learning Research, vol. 80, PMLR, Oct. 2018, pp. 5502–5511. [Online]. Available: `https://proceedings.mlr.press/v80/xu18h.html`.

[141] H. Al-Behadili, K. Ku-Mahamud, and R. Sagban, "Rule pruning techniques in the antminer classification algorithm and its variants: A review," Apr. 2018. DOI: `10.1109/ISCAIE.2018.8405448`.

[142] R. Senge, S. Bösner, K. Dembczyński, J. Haasenritter, O. Hirsch, N. Donner-Banzhoff, and E. Hüllermeier, "Reliable classification: Learning classifiers that distinguish aleatoric and epistemic uncertainty," *Information Sciences*, vol. 255, pp. 16–29, Jan. 2014. DOI: `10.1016/j.ins.2013.07.030`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0020025513005410`.

[143] J. Shaw, F. Rudzicz, T. Jamieson, and A. Goldfarb, "Artificial intelligence and the implementation challenge," *J Med Internet Res*, vol. 21, no. 7, e13659, Jul. 2019, ISSN: 1438-8871. DOI: `10.2196/13659`. [Online]. Available: `http://www.ncbi.nlm.nih.gov/pubmed/31293245`.

[144] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44–56, Jan. 2019. DOI: `10.1038/s41591-018-0300-7`. [Online]. Available: `https://www.nature.com/articles/s41591-018-0300-7#citeas`.

[145] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 11–29, 2020. DOI: `10.1109/JPROC.2019.2932116`.

[146] F. C. Ghesu, E. Krubasik, B. Georgescu, V. Singh, Y. Zheng, J. Hornegger, and D. Co-maniciu, "Marginal space deep learning: Efficient architecture for volumetric image parsing," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1217–1228, May 2016. DOI: 10.1109/tmi.2016.2538802. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7426845.

[147] F. Pereira, A. Bueno, A. Rodriguez, D. Perrin, G. Marx, M. Cardinale, I. Salgo, and P. del Nido, "Automated detection of coarctation of aorta in neonates from two-dimensional echocardiograms," *Journal of Medical Imaging*, vol. 4, no. 1, p. 014502, Jan. 2017. DOI: 10.1117/1.jmi.4.1.014502. [Online]. Available: https://www.spiedigitallibrary.org/journals/journal-of-medical-imaging/volume-4/issue-1/014502/Automated-detection-of-coarctation-of-aorta-in-neonates-from-two/10.1117/1.JMI.4.1.014502.full?SSO=1.

[148] M. Xian, Y. Zhang, H. Cheng, F. Xu, B. Zhang, and J. Ding, "Automatic breast ultra-sound image segmentation: A survey," *Pattern Recognition*, vol. 79, pp. 340–355, Jul. 2018. DOI: 10.1016/j.patcog.2018.02.012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320318300645.

[149] D. Lyell and E. Coiera, "Automation bias and verification complexity: a systematic re-view," *Journal of the American Medical Informatics Association*, vol. 24, no. 2, pp. 423–431, Aug. 2016, ISSN: 1067-5027. DOI: 10.1093/jamia/ocw105. eprint: https://academic.oup.com/jamia/article-pdf/24/2/423/34148461/ocw105.pdf. [Online]. Available: https://doi.org/10.1093/jamia/ocw105.

[150] F. Légaré, A. M. O'Connor, I. D. Graham, D. Saucier, L. Côté, J. Blais, M. Cauchon, and L. Paré, "Primary health care professionals' views on barriers and facilitators to the implementation of the ottawa decision support framework in practice," *Patient Education and Counseling*, vol. 63, no. 3, pp. 380–390, 2006, 3rd International Conference on Shared Decision Making, ISSN: 0738-3991. DOI: https://doi.org/10.1016/j.pec.2006.04.011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0738399106001315.

[151] J. Merkert, M. Mueller, and M. Hubl, "A survey of the application of machine learning in decision support systems.," in *ECIS*, 2015.

[152] A. Awaysheh, J. Wilcke, F. Elvinger, L. Rees, W. Fan, and K. L. Zimmerman, "Review of medical decision support and machine-learning methods," *Veterinary pathology*, vol. 56, no. 4, pp. 512–525, 2019.

[153] E. Vayena, A. Blasimme, and I. G. Cohen, "Machine learning in medicine: Addressing ethical challenges," *PLoS medicine*, vol. 15, no. 11, e1002689, 2018.

[154] A. Ng and J. Swanevelder, "Resolution in ultrasound imaging," *Continuing Education in Anaesthesia Critical Care Pain*, vol. 11, no. 5, pp. 186–192, Aug. 2011, ISSN: 1743-1816. DOI: 10.1093/bjaceaccp/mkr030. eprint: https://academic.oup.com/bjaed/article-pdf/11/5/186/794418/mkr030.pdf. [Online]. Available: https://doi.org/10.1093/bjaceaccp/mkr030.

[155] E. W. Prince, R. Whelan, D. M. Mirsky, N. Stence, S. Staulcup, P. Klimo, R. C. E. Anderson, T. N. Niazi, G. Grant, M. Souweidane, J. M. Johnston, E. M. Jackson, D. D. Limbrick, A. Smith, A. Drapeau, J. J. Chern, L. Kilburn, K. Ginn, R. Naftel, and R. Dudley, "Robust deep learning classification of adamantinomatous craniopharyngioma from limited preoperative radiographic images," *Scientific Reports*, vol. 10, no. 1, Oct. 2020. DOI: 10.1038/s41598-020-73278-8. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7547020/.

[156] K. Hu, D. Liu, S. Herrmann, M. Niemann, P. D. Gaudron, W. Voelker, G. Ertl, B. Bijnens, and F. Weidemann, "Clinical implication of mitral annular plane systolic excursion for patients with cardiovascular disease," *European Heart Journal - Cardiovascular Imaging*, vol. 14, no. 3, pp. 205–212, Nov. 2012.

[157] M. C. ELISH, "The stakes of uncertainty: Developing and integrating machine learning in clinical care," *Ethnographic Praxis in Industry Conference Proceedings*, vol. 2018, no. 1, pp. 364–380, Oct. 2018. DOI: 10.1111/1559-8918.2018.01213. [Online]. Available: https://anthrosource.onlinelibrary.wiley.com/doi/epdf/10.1111/1559-8918.2018.01213.

[158] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "Xai—explainable artificial intelligence," *Science Robotics*, vol. 4, no. 37, eaay7120, 2019.

[159] E. de Souza Nascimento, I. Ahmed, E. Oliveira, M. P. Palheta, I. Steinmacher, and T. Conte, "Understanding development process of machine learning systems: Challenges and solutions," in *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, IEEE, 2019, pp. 1–6.

[160] Y. H. Yoon, S. Khan, J. Huh, and J. C. Ye, "Efficient b-mode ultrasound image reconstruction from sub-sampled rf data using deep learning," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 325–336, 2018.

[161] K. Shaheen, M. A. Hanif, O. Hasan, and M. Shafique, "Continual learning for real-world autonomous systems: Algorithms, challenges and frameworks," *Journal of Intelligent & Robotic Systems*, vol. 105, no. 1, pp. 1–32, 2022.

[162] B. Liu, "Learning on the job: Online lifelong and continual learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 13 544–13 549.

[163] T. Grote and P. Berens, "How competitors become collaborators—bridging the gap(s) between machine learning algorithms and clinicians," *Bioethics*, vol. 36, no. 2, pp. 134–142, Oct. 2021. DOI: 10.1111/bioe.12957. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1111/bioe.12957.

[164] A. Kiani, B. Uyumazturk, P. Rajpurkar, A. Wang, R. Gao, E. Jones, Y. Yu, C. P. Langlotz, R. L. Ball, T. J. Montine, B. A. Martin, G. J. Berry, M. G. Ozawa, F. K. Hazard, R. A. Brown, S. B. Chen, M. Wood, L. S. Allard, L. Ylagan, and A. Y. Ng, "Impact of a deep learning assistant on the histopathologic classification of liver cancer," *npj Digital Medicine*, vol. 3, no. 1, Feb. 2020. DOI: 10.1038/s41746-020-0232-8. [Online]. Available: https://www.nature.com/articles/s41746-020-0232-8#citeas.

[165]   M. Grissinger, "Understanding human over-reliance on technology," *Pharmacy and Therapeutics*, vol. 44, no. 6, p. 320, 2019.

[166]   U. Food, D. Administration, *et al.*, "Developing a software precertification program: A working model," *US Department of Health and Human Services*, 2018.

[167]   H. Fan, L. Lin, F. Yang, P. Chu, G. Deng, S. Yu, H. Bai, Y. Xu, C. Liao, and H. Ling, *Lasot: A high-quality benchmark for large-scale single object tracking*, 2018. [Online]. Available: `https://arxiv.org/abs/1809.07845`.

[168]   M. Müller, A. Bibi, S. Giancola, S. Al-Subaihi, and B. Ghanem, *Trackingnet: A large-scale dataset and benchmark for object tracking in the wild*, 2018. [Online]. Available: `https://arxiv.org/abs/1803.10794`.

[169]   L. Huang, X. Zhao, and K. Huang, "Got-10k: A large high-diversity benchmark for generic object tracking in the wild," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 5, pp. 1562–1577, 2021. DOI: `10.1109/TPAMI.2019.2957464`.

[170]   M. Mueller, N. Smith, and B. Ghanem, *A benchmark and simulator for uav tracking*, Oct. 2016. [Online]. Available: `https://www.researchgate.net/publication/308278377_A_Benchmark_and_Simulator_for_UAV_Tracking`.

[171]   H. K. Galoogahi, A. Fagg, C. Huang, D. Ramanan, and S. Lucey, *Need for speed: A benchmark for higher frame rate object tracking*, 2017. [Online]. Available: `https://arxiv.org/abs/1703.05884`.

[172]   Y. Wu, J. Lim, and M.-H. Yang, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834–1848, Sep. 2015. DOI: `10.1109/tpami.2014.2388226`. [Online]. Available: `https://pubmed.ncbi.nlm.nih.gov/26353130/`.

[173]   M. Kristan, A. Leonardis, J. Matas, M. Felsberg, R. Pflugfelder, L. Cehovin Zajc, T. Vojir, G. Bhat, A. Lukezic, A. Eldesokey, and et al., "The sixth visual object tracking vot2018 challenge results," *Thecvf.com*, 2018. [Online]. Available: `https://openaccess.thecvf.com/content_eccv_2018_workshops/w1/html/Kristan_The_sixth_Visual_Object_Tracking_VOT2018_challenge_results_ECCVW_2018_paper.html`.

[174]   594422814, *Github - 594422814/transformertrack*, May 2021. [Online]. Available: `https://github.com/594422814/TransformerTrack`.

[175]   594422814, *594422814/transformertrack*, May 2021. [Online]. Available: `https://github.com/594422814/TransformerTrack`.

[176]   2021. [Online]. Available: `https://docs.conda.io/en/latest/`.

[177]   JetBrains, *Pycharm*, 2021. [Online]. Available: `https://www.jetbrains.com/pycharm/`.

[178]   visionml, *Github - visionml/pytracking: Visual tracking library based on pytorch*. Nov. 2021. [Online]. Available: `https://github.com/visionml/pytracking`.

[179]   pytorch, *Github - pytorch/pytorch: Tensors and dynamic neural networks in python with strong gpu acceleration*, Feb. 2022. [Online]. Available: `https://github.com/pytorch/pytorch#more-about-pytorch`.

[180]   2022. [Online]. Available: `https://www.h5py.org/`.

[181] M. McKendrick, S. Yang, and G. A. McLeod, "The use of artificial intelligence and robotics in regional anaesthesia," *Anaesthesia*, vol. 76, no. S1, pp. 171–181, Jan. 2021. DOI: 10.1111/anae.15274.

Appendix A

Technical Commands

Use an X Window System. XQuartz is suitible for Mac OS. The commands are the ones used in this thesis. For further use some modifications to the directories and usernames must be done.

A.1 Training

First, log in to the server with the given username and type the correct password.

```
$ ssh -lvwoien -Y idi.bohaga.com
```

To edit the code in PyCharm, open the editor with:

```
$ pycharm-community
```

Before training, activate the conda environment previously generated. Here, the environment is called *pytracking*.

```
$ conda activate pytracking
```

To train on the server:

```
$ nohup python run_training.py dimp transformer_dimp > nuhup.out &
```

The output of the training will appear in the nohup.out file, which can be named whatever.

The network checkpoints is saved:
"$/home/vwoien/Transf ormerTrack/ltr/network\_checkpoints$"
This path can be changed in ltr/admin/local.py.

In order to see what is running on the server:

```
$ nvidia-smi
```

To show all the GPU activity:

```
$ ps -aux
```

To end the training, before it terminates. Type in the correct process id instead of "pid".

```
$ kill "pid"
```

A.2   Vizualising the Results from Training

A.2.1   Visalize with TensorBoard

In order to plot the results locally on a personal computer, copy the results to e.g.. the desktop:

```
$ scp -r  vwoien@idi.bohaga.com:path_of_files_to_copy path_of_new_local_location
```

scp = secure copy, -r = recursive

In order to vizualize the results with TensorBoard:

```
$ pip install tensorboard
```

Then:

```
$ tensorboard --logdir Desktop/transformer_dimp (PATH)
```

Serving TensorBoard on localhost: http://localhost:6006/.

A.3   Tracking

The local paths for dataset and network used for tracking is set in:
$/home/vwoien/TransformerTrack/pytracking/evaluation/local.py$

The pre-trained network is saved in:
$/home/vwoien/TransformerTrack/pytracking/networks/trdimp\_net.pth.tar$

All networks are saved in:
$/home/vwoien/TransformerTrack/pytracking/networks/$
Run the trdimp tracker:

```
$ python run_tracker.py trdimp trdimp --dataset_name ultrasound_test
```

Track using nohup:

```
$ nohup python run_tracker.py trdimp trdimp --dataset_name ultrasound_test > nuhup.out &
```

The tracking results is saved in:
$/home/vwoien/TransformerTrack/pytracking/tracking\_results$
This path can be changed in $/home/vwoien/TransformerTrack/pytracking/evaluation/local.py$

A.4 Vizualising the Results from Testing

Vizualize the tracking results and calculate the qualitative tracking performance by running:
/home/vwoien/TransformerTrack/pytracking/results_imgs/plot_results_imgs.py

The images are saved in a folder with the respective dataset name, and the qualitative tracking performance measure is saved in a txt file inside the corresponding folder here:
/home/vwoien/TransformerTrack/pytracking/results_imgs/

To generate and save a video of all the frames with the visualization of the tracking performance, run this file:
/home/vwoien/TransformerTrack/pytracking/results_vids/generate_video_from_imgs.py

The videos are saved in a folder with the respective dataset name in:
/home/vwoien/TransformerTrack/pytracking/results_vids/

A.5 Programs for Data Reprocessing

To annotate data: */home/vwoien/TransformerTrack/ltr/data/annotate.py*
To transform the data in the correct format: */home/vwoien/TransformerTrack/ltr/data/data_loader.py*
To split the data in train, test, and validation set:
/home/vwoien/TransformerTrack/ltr/data/train_test_val_split.py

A.6 Other Terminal Commands

Here is an overview of other terminal commands that may be useful when training and testing via the server.

Show text files in the terminal (change text.txt):

```
$ cat text.txt
```

Make a directory:

```
$ mkdir name_of_dir
```

Print the working directory:

```
$ pwd
```

Prints out files and directories that are existing in the current directory (or write another directory after ls):

```
$ ls
```

Change directory (directory is e.g. home/desktop/data):

```
$ cd directory
```

Appendix B

Libraries and Packages

The TransformerTrack method developed by Wang et al. [49] is open-source [174]. The repository, submodules and dependencies can be cloned from GitHub [175]. The requirements for the repository are Conda, an open-source package- and environment management system, installation with Python 3.7 and Nvidia graphics processing unit (GPU) [49, 176].

TransformerTrack is coded in Python. PyCharm was used in this paper, a popular integrated development environment for Python development [177]. TransformerTrack is based on Py-Tracking, a general Python framework for visual object tracking and video object segmentation [178]. The framework is favorited by 2000 users at GitHub, indicating that developers like the framework. The framework include an official implementation of the DiMP tracker, which is the tracker used in this paper. PyTracking is based on PyTorch, an open source machine learning framework [179]. PyTorch is widely used, and 54200 GitHub users have favorited the code at GitHub.

HDFView was used to view the h5 files. A Python interface was used to open and read the h5 files in Python, called the h5py package [180]. The *Python Imaging Library (PIL)* was used for image processing operations, such as reading and saving images.

Tensorflow was used to generate the loss plot for evaluating the training performance.

The flowcharts were made using the online diagram softwares *Lucidchart* and *draw.io*.

Appendix C

Concepts of Artificial Intelligence

Concept	Definition
Artificial intelligence	A non-human programme or model that can solve sophisticated tasks
Machine learning	A subset of artificial intelligence, characterised by improvements in performance through iterative tuning of weights or coefficients within mathematical models.
Deep learning	A multi-layered, non-linear extension of machine learning inspired by neural networks in the brain.
Bounding box	In an image, the (x, y) co-ordinates of a rectangle around an area of interest.
Convolutional network	A neural network typically consisting of convolutional, pooling and dense (fully connected) layers.
Intersection over union	Measures the accuracy of the model's predicted bounding box with respect to the ground-truth bounding box. It is the ratio between the overlapping area and the total area.
Overfitting	A model that matches the training data so closely that the model fails to make correct predictions on new data.
Underfitting	A model with poor predictive ability that has failed to capture the complexity of the training data.
Pooling	Reducing a matrix created by an earlier convolutional layer to a smaller matrix by taking the maximum or average value.
Reinforcement learning	A family of algorithms that maximise return.
Supervised learning	Training a model from input data and their corresponding labels.
Test set	The subset of the dataset used to test the model after validation.
Training set	The subset of the dataset used to train a model.
Validation set	A subset of the dataset, disjoint from the training set, used in validation.

Table C.1: Some concepts of AI that is relevant for ultrasound object tracking. The definitions are directly from McKendrick et. al [181].

Appendix D

Interview

Overview of the questions the interviews were based on:

1. This autumn, I mapped how the processes in an operating room take place and how they will be after implementing the digital monitoring tool. I would like us to look at those flow charts together. Can you please provide corrections or questions to the mappings? (*Show the As-Is and To-Be process mappings for the Specialization Project thesis*)
2. How and when is the threshold value sat?
3. Is the threshold value different for each patient?
4. What must a doctor think about when deciding the threshold value?
5. When deciding the threshold value, do doctors think about any problems with false positive and false negative outputs?
6. When using the new tool, is there anything the ultrasound images can inform the doctors about to make the operation take longer or require more resources?
7. Are there any concerns about a machine learning model being trained only on ill hearts?
8. How are you experiencing using the new tool?