

Juulia Saarinen

ATLAS-BASED TEMPLATE MATCHING TO DETERMINE KNEE ALIGNMENT FROM PLAIN LONG-LEG RADIOGRAPHS

Master of Science Thesis
Faculty of Medicine and Health Technology
Examiners: Prof. Mark van Gils
M.Sc. Antti Kallonen
Dr. Antti Eskelinen
June 2025

ABSTRACT

Juulia Saarinen: Atlas-based template matching to determine knee alignment from plain long-leg radiographs

Master of Science Thesis

Tampere University

Master's Programme in Biotechnology and Biomedical Engineering

June 2025

This Master of Science thesis was conducted as a commissioned study for Coxa Hospital for Joint Replacement, a hospital specializing in joint replacement surgeries. The majority of its procedures are knee replacement surgeries. Knee osteoarthritis is the most common condition leading to knee replacement, and angular deformities of the knee accelerate the progression of osteoarthritis. Angular deformities of the knee are classified into varus and valgus types. In varus deformity, the knee bends laterally, and in valgus deformity, the knee bends medially. These must be corrected during knee joint replacement surgery to achieve long-term results.

Determining lower limb alignment from a plain long-leg radiograph is an important part of planning knee replacement surgery and assessing the postoperative alignment. Lower limb alignment is defined using the hip-knee-ankle angle. To determine this angle, the center of the femoral head, the center of the knee joint, and the center of the distal end of the tibia must be accurately marked on the radiographic image. The hip-knee-ankle angle is calculated based on the angle between the lines connecting the ankle to the knee and the knee to the hip. However, manual annotation is prone to error. Different orthopedic surgeons may mark slightly different locations, and even the same surgeon may mark the points differently on different days. Even a small difference in these markings can lead to a significant error in alignment measurement, which in turn affects the success of the knee replacement surgery and the longevity of the prosthesis.

The aim of this thesis is to design, implement, and evaluate a model based on traditional computer vision methods that determines the angular deformity of the knee and measures the hip-knee-ankle angle from long-leg radiographic images. The radiographs were preprocessed to improve image quality and to make joint detection easier. Histogram equalization was used for preprocessing. The hip, knee, and ankle joints were localized using an atlas-based template matching. The hip-knee-ankle angle was measured based on the centers of the matched templates, and the results were compared to manual measurements made by an experienced orthopedic surgeon. Two different methods were used to determine angular deformity. In the first method, the position of the localized knee joint was compared to the mechanical axis of the lower limb defined by the localized hip and ankle joints. In the second method, the position of the knee joint was compared directly to the localized ankle joint.

The automatic measurements matched moderately with the manual measurements. The mean absolute error between orthopedists' and algorithm measurements was 4.5 degrees, and the intraclass correlation coefficient was 0.76. Better accuracy in determining the type of angular deformity was achieved when comparing the relative position of the knee joint to the ankle joint. Challenges in localizing the femoral head impaired the definition of the mechanical axis and therefore affected the first method. The femoral head was the most difficult joint to locate, partly due to its position inside the pelvis and partly because of potential existing hip prostheses. As a result, more templates were needed for the hip atlases than for the other joints. In the future, locating joints based on known joint coordinates instead of only hip-knee-ankle angle values could bring more reliable results. Deep learning could also be utilized to enhance and refine analysis.

Keywords: image processing, knee alignment, hip-knee-ankle angle, osteoarthritis, radiograph, template matching, computer vision

The originality of this thesis has been checked using the Turnitin Originality Check service.

TIIVISTELMÄ

Juulia Saarinen: Atlas-pohjainen mallin sovitus määrittämään alaraajan linjauksen röntgenkuvasta

Diplomityö

Tampereen yliopisto

Bioteknologian ja biolääketieteen tekniikan maisteriohjelma

Kesäkuu 2025

Tämä diplomityö suoritettiin toimeksiantona Tekonivelsairaala Coxa Oy:lle, joka on tekonivelleikkauksiin erikoistunut sairaala. Suurin osa Coxan leikkauksista on polven tekonivelleikkauksia. Polven nivelrikko on yleisin polven tekonivelleikkaukseen johtava sairaus, ja polven virheasento nopeuttaa polven nivelrikon etenemistä. Polven virheasennot jaetaan varus- ja valgusasentoihin. Varusasennossa polvi taittuu lateraalisuuntaan, ja valgusasennossa polvi taittuu mediaalisuuntaan. Virheasento tulee korjata polven tekonivelleikkauksessa pitkäaikaisten tulosten saavuttamiseksi.

Alaraajan linjauksen määrittäminen röntgenkuvasta on tärkeä osa polven tekonivelleikkauksen suunnittelua ja leikkauksen jälkeistä polven asennon arvioimista. Alaraajan linjaus määritetään niin kutsutun lonkka-polvi-nilkka kulman (hip-knee-ankle angle) avulla. Tämän määrittämiseksi alaraajan röntgenkuvaan tulee merkitä tarkasti reisiluun pään keskikohta, polvinivelen keskikohta, sekä sääriluun distaalisen pään keskikohta. Kulma määritetään nilkan ja polven, sekä polven ja lonkan välisten janojen kulmana. Manuaalinen merkitseminen on kuitenkin herkkä virheille. Ortopedit voivat asettaa merkit hiukan eri paikkoihin, ja eroja voi olla myös saman ortopedin eri päivinä tekemissä merkinnöissä. Pienikin ero kohtien asettamisessa voi johtaa suureen eroon linjauksen määrittämisessä, ja vaikuttaa siten polven tekonivelleikkauksen onnistumiseen ja tekonivelen kestävyYTEEN.

Työn tarkoituksena on suunnitella, toteuttaa, ja arvioida perinteiseen tietokonenäköön pohjautuva malli, joka määrittää polven virheasennon ja mittaa alaraajan linjauksen alaraajan röntgenkuvista. Röntgenkuvia esikäsiteltiin kuvan laadun parantamiseksi ja nivelkohtien löytämisen helpottamiseksi. Esikäsitelyä käytettiin histogrammin tasoitusta. Lonkka-, polvi-, ja nilkkanivel paikallistettiin atlas-pohjaiseen mallin sovitukseen avulla. Alaraajan linjaus mitattiin perustuen paikallistettujen mallien keskikohtiin, ja mitattuja arvoja verrattiin kokeneen ortopedin tekemiin manuaalisiin mittauksiin. Virheasennon määrittämiseen käytettiin kahta eri menetelmää. Ensimmäisessä paikallistettua polvinivelen sijaintia verrattiin paikallistettujen lonkan ja nilkan avulla määritettyyn alaraajan mekaaniseen akseliin, ja toisessa menetelmässä paikallistettua polvinivelen sijaintia verrattiin paikallistettuun nilkan sijaintiin.

Automaattiset mittaukset vastasivat manuaalisia mittauksia kohtalaisella tarkkuudella. Keskimääräinen absoluuttinen virhe ortopedin ja algoritmin mittauksen välillä on 4,5 astetta, ja ICC-arvo oli 0,76. Parempi tarkkuus virheasennon tyyppin määrittämiseen saatiin vertaamalla polvinivelen sijaintia suhteessa nilkan sijaintiin. Mekaanisen akselin määrittämistä ja siten ensimmäistä vaihtoehtoa häiritsi reisiluun pään paikallistamisen haasteet. Reisiluun pää oli nivelkohdista kaikista vaikein paikantaa, johtuen sen sijainnista osittain lonkkaluun sisällä, sekä mahdollisista kuvissa näkyvistä lonkkaproteeseista. Näin ollen myös lonkan atlatset tarvitsivat eniten malleja muihin niveliin verrattuna. Jatkossa nivelten paikantaminen tunnettujen nivelkoordinaattien perusteella pelkkien tunnettujen lonkka-polvi-nilkka kulmien sijaan toisi luotettavampia tuloksia. Syväoppimista voitaisiin myös hyödyntää analyysin tehostamiseen ja tarkentamiseen.

Avainsanat: kuvankäsittely, alaraajan linjaus, nivelrikko, röntgenkuva, mallin sovitus, tietokonenäkö

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin Originality Check -ohjelmalla.

USE OF AI IN THESIS

I have utilised AI tools in my thesis:

- No
- Yes

The AI tools utilised in my thesis and their purposes are described below:

Names and versions of AI tools: ChatGPT (GPT-4)

Purpose of using AI tools: To correct grammatical errors, structure the text, and assist in code creation.

Sections where AI tools were used: Throughout the thesis.

I acknowledge that I am fully responsible for the entire content of my thesis, including the parts generated by AI, and accept accountability for any violations of ethical standards in publications.

PREFACE

I want to express my gratitude to Coxa Hospital for Joint Replacement, and especially to the Head of Research and Innovation, Dr. Antti Eskelinen, and the Head of Research, Dr. Aleksi Reito, for providing a fascinating thesis topic that is both clinically significant and highly relevant in today's context. I'm thankful for the opportunity to challenge myself and acquire new knowledge and skills through this work.

I am grateful for the guidance I received throughout the process. My thanks go to my supervisors, Professor Mark van Gils and M.Sc. Antti Kallonen, for their mentorship, support, insightful ideas, and constructive feedback. Their guidance gave me confidence in my work and helped me move forward whenever I faced obstacles.

I'm especially thankful to my family and friends, whose unwavering support throughout my studies, particularly during the thesis process, meant the world to me. Their encouragement helped me push my limits and stay motivated to complete my studies with integrity and dedication, while also reminding me to make time for joy and laughter along the way.

Finally, I want to thank my fiancé, Waldemar, for his endless support and belief in me – not only in completing this thesis, but in everything I choose to pursue in life.

Tampere, 9th June 2025

Juulia Saarinen

CONTENTS

1. Introduction	1
1.1 Objectives of the study	2
1.2 Structure of the thesis	3
2. Background.....	4
2.1 Medical context.....	4
2.1.1 Lower limb alignment	4
2.1.2 Osteoarthritis	6
2.1.3 Total knee arthroplasty.....	8
2.1.4 Radiographic imaging	9
2.1.5 Coxa Hospital for Joint Replacement	11
2.2 Medical image processing	12
2.2.1 Computer vision in medical image processing	14
2.2.2 Automation of radiographic analysis.....	15
2.3 Techniques to determine hip-knee-ankle angle.....	17
2.3.1 Manual and semi-automated methods	17
2.3.2 Computer vision-based methods.....	17
3. Materials And Methods	21
3.1 Dataset.....	21
3.2 Data processing.....	22
3.3 Manual measurement procedure and narrowing the dataset	23
3.4 Image preprocessing	24
3.4.1 Noise reduction.....	24
3.4.2 Histogram equalization.....	25
3.4.3 Edge detection.....	27
3.5 Template matching	28
3.6 Statistical analysis	33
3.6.1 Error.....	33
3.6.2 Correlation	34
4. Results.....	36
4.1 Template matching	36
4.2 Statistical analysis	38
4.3 Graphical analysis	39
4.4 Visual review of images	42
5. Discussion	45
6. Conclusions	50
References.....	51

LIST OF ABBREVIATIONS

AA	Anatomic axis
AI	Artificial intelligence
BMI	Body mass index
CAD	Computer-aided diagnosis
CNN	Convolutional neural network
CT	Computer tomography
CV	Computer vision
DICOM	Digital Imaging and Communications in Medicine
DIP	Digital image processing
DL	Deep learning
FTA	Femorotibial angle
HKAA	Hip-knee-ankle angle
ICC	Intraclass correlation coefficient
KL	Kellgren-Lawrence
LLR	Long-leg radiograph
MA	Mechanical axis
MAD	Mechanical axis deviation
MEA	Mean average error
MI	Medical imaging
MIP	Medical image processing
MRI	Magnetic resonance imaging
OA	Osteoarthritis
PCC	Pearson correlation coefficient
R-CNN	Region-based convolutional neural networks
RMSE	Root mean square error
TKA	Total knee arthroplasty

1. INTRODUCTION

Determining lower limb alignment is an important part of planning, performing, and ensuring the long-term success of total knee arthroplasty (TKA). The gold standard for this assessment is measuring the hip-knee-ankle angle (HKAA) (Zorzi & Miranda, 2018). According to a study by the Finnish Institute for Health and Welfare (THL, 2024), knee arthroplasty is the most common joint replacement surgery in Finland. In 2023, a total of 14,716 primary knee arthroplasties were performed in Finland, accounting for 54% of all primary joint replacements (hip, knee, and shoulder) that year (THL, 2024). Due to the aging population, limited resources in joint replacement surgery, and the backlog of elective procedures from the COVID-19 period in 2020-2022, the need for knee replacements continues to grow (Tapio, 2025).

Coxa Hospital for Joint Replacement (Coxa) is the largest hospital in Finland and in the Nordic countries specializing in joint replacement surgery (Coxa, 2024). In 2023, over 25% of all knee replacements in Finland (4101 in total; 3914 primary and 187 revision surgeries) were performed at Coxa (THL, 2024). The most common cause of knee arthroplasty is osteoarthritis, for which knee replacement is the only curative treatment. In 2023, 95% of the knees operated in Coxa and elsewhere in Finland had osteoarthritis (THL, 2024).

Lower limb alignment is typically assessed using long-leg radiographs (LLRs), which show the entire lower limb from the femoral head to the ankle joint in a single image (Zorzi & Miranda, 2018). The HKAA is manually measured by an orthopedic surgeon by placing points on the X-ray at the femoral head, center of the knee joint, and center of the ankle joint (Marques Luís & Varatojo, 2021). However, this process is slow, laborious, and prone to human error. Even small variations in point placement can lead to significant differences in the HKAA, which can result in suboptimal implant positioning and increase the likelihood of early revision surgery. These variations occur between different orthopedic surgeons and can depend on the surgeon's experience or level of alertness. Each surgeon may place the points slightly differently, leading to lower reproducibility of the measurements.

The growing number of knee replacements and the challenges of clearly defining lower limb alignment with current methods have created a need for reliable, independent, objective, and fast measurements to support clinical decision-making. Computer-assisted methods have become more common in medical imaging during this century. In particular, artificial intelligence has recently gained popularity in medical image segmentation, classification, and registration (Elyan et al., 2022). However, problems include small dataset sizes and the use of highly similar images for training the algorithm.

This thesis aims to develop a template-matching-based computer-vision algorithm to accurately measure hip-knee-ankle angle and to detect alignment type from plain LLRs. The thesis is conducted as a commissioned study for Coxa Hospital for Joint Replacement.

1.1 Objectives of the study

The primary aim of this study is to develop an image processing pipeline that accurately identifies lower limb bone and joint surfaces, measures hip-knee-ankle angle and determines lower limb deformity type from plain radiographs. Algorithm measurements are compared to an orthopedist's measurements. This HKAA measurement pipeline is described in Figure 1.1.

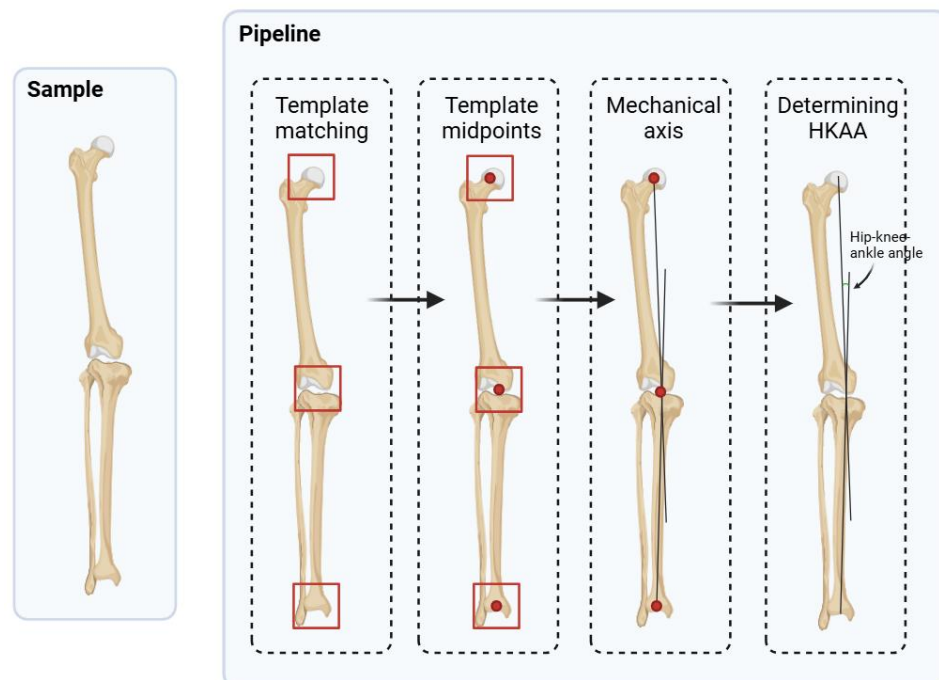


Figure 1.1. Visual representation of the hip-knee-ankle angle measurement pipeline. First, the hip, knee, and ankle joints are localized using template matching. Next, the midpoint of each matched template is determined. Using these midpoints, the femoral and tibial mechanical axes are constructed, and the HKAA is calculated using the laws of cosine. Created with BioRender.com

The aim is to provide consistent, repeatable, and reliable HKAA measurements. The goal is to reduce inter- and intra-observer variability between measurements by providing a tool that offers standardized measurements across different clinical settings and among different clinicians, thereby improving the overall reliability of knee alignment assessment. In addition, the study aims to explore image processing methods that best highlight joint locations in radiographs. The research questions are listed below:

1. How closely do the algorithm's measurements align with the orthopedist's measurements?
2. How accurately does the algorithm detect angular deformity?
3. How many templates are needed to receive reliable results?
4. Which joint is the easiest or hardest to locate in radiographic images?

This work also aims to identify the main challenges in developing the HKAA measurement algorithm and to consider the characteristics the dataset should have in order to build a reliable algorithm suitable for use in a clinical setting.

1.2 Structure of the thesis

First, this thesis provides an understanding of the background relevant to the study. It includes introducing the concept of knee alignment and relevant medical background, as well as different technical methods to process medical images and existing methods to determine hip-knee-ankle angle from plain radiographs. After that, the dataset and the image analysis methods used in the study are described in detail before presenting the results. Finally, the implications of the findings are discussed, and conclusions are drawn, along with recommendations for future research.

2. BACKGROUND

The background section is divided into three main topics: The medical context, which covers medical definitions, conditions, and limitations related to lower limb alignment and its assessment; the principles and goals of processing medical images, particularly radiographs; and the currently used and emerging methods for assessing lower limb alignment.

2.1 Medical context

This section provides a comprehensive overview of lower limb alignment assessment, disorders caused by misalignment, such as knee osteoarthritis, and the importance of proper alignment in the planning and success of knee replacement surgery. Medical context establishes the need for automatic alignment assessment and guides the automation to meet real-world clinical requirements, considering the limitations and challenges associated with radiographic imaging. Understanding the medical context and the origin of the images is essential, as these factors largely determine the direction of the automation and define what kind of data is available and usable, as well as how precise the results can be. The chapter also introduces Coxa Hospital for Joint Replacement, its role in joint replacement care in Finland, its current methods for assessing knee alignment, and its efforts to develop automated solutions.

2.1.1 Lower limb alignment

Lower limb alignment refers to how the bones and joints of the leg – the femur, tibia, and knee joint – are positioned relative to each other and to the body's overall posture (Marques Luís & Varatojo, 2021). Proper alignment is critical for maintaining balance, efficient movement, and joint health. Misalignment can lead to pain, improper weight distribution, and long-term joint problems, such as osteoarthritis. Variables like hip-knee-ankle angle (HKAA), mechanical axis deviation (MAD), and femorotibial angle (FTA) are used to assess lower limb alignment. HKAA is determined as an angle between femoral mechanical axis (a line from the center of the femoral head to the center of knee joint) and tibial mechanical axis (a line from the center of knee joint to the center of tibial plafond) (Figure 2.1A) (Jia et al., 2023). The mechanical axis deviation is defined as the perpendicular distance between the knee center and the mechanical axis (MA) of the lower limb (a line joining femur and tibia center) (Figure 2.1B) (Insall et al., 1985). The femorotibial angle is the angle between the femoral and tibial anatomical axes (AA) of

the lower limb (Figure 2.1C). The anatomical axis is a line centered in the diaphysis of each bone. Determining the HKAA, MAD, and FTA is illustrated in Figure 2.1. (Marques Luís & Varatojo, 2021)

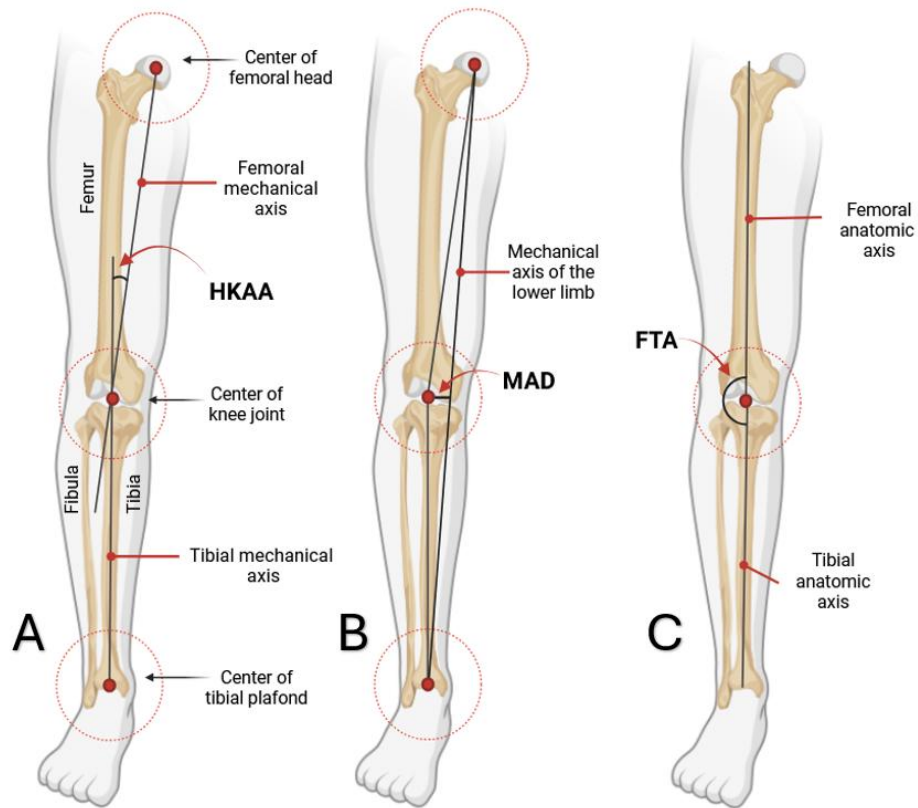


Figure 2.1. Determining the hip-knee-ankle angle (A), mechanical axis deviation (B), and femorotibial angle (C). Created with BioRender.com, adapted from Marques Luís & Varatojo (2021).

Despite all the alignment determination methods, HKAA is used as a gold standard when determining lower limb alignment since it allows for consistent and precise measurement of mechanical tibiofemoral angle and assessment of limb deformities (Marques Luís & Varatojo, 2021). In addition to that, other determination methods have their limits. According to Choudhury et al. (2023), MAD is affected by an individual's height. Taller individuals tend to have increased MAD compared to shorter individuals with the same hip-knee-ankle angle, making MAD less suitable for general use (Choudhury et al., 2023). FTA is often used to predict the HKAA, although AA length influences the ability to estimate HKAA. Correlation between FTA and HKAA is much weaker for shorter-shaft FTA measurement (Sheehy et al., 2011), causing imprecision when using in operative planning (Sheehy & Cooke, 2015). Therefore, it is clinically most relevant to use HKAA to assess lower limb alignment.

In normal knee alignment, the mechanical axis of the lower limb runs from the center of the hip joint, through the center of the knee joint, to the center of the ankle joint. Knee

alignment is divided into varus and valgus deformation depending on whether the lower limb axis deviates medially (varus) or laterally (valgus) (Figure 2.2) (Cooke et al., 1997).

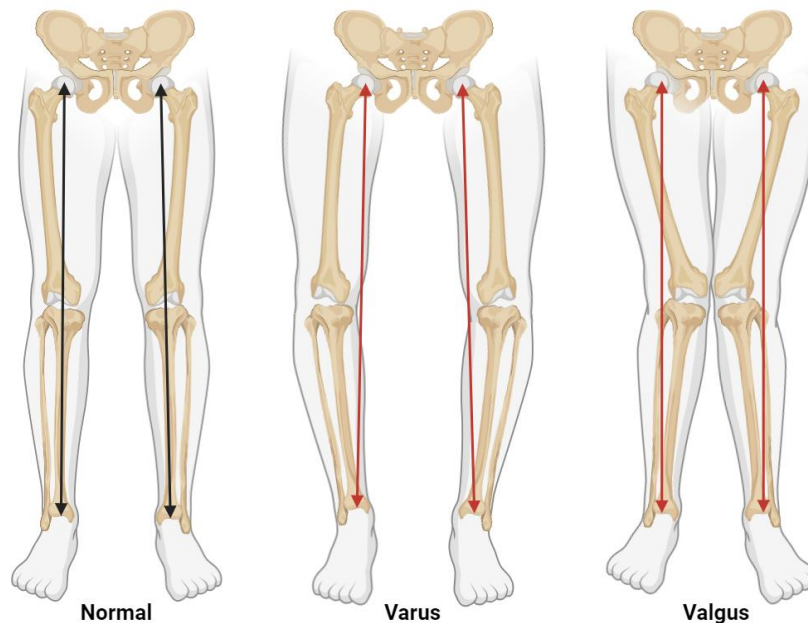


Figure 2.2. *Varus and valgus deformities. The normal mechanical axis of the lower limb passes from the center of the femoral head through the knee joint to the center of the ankle joint. In varus deformation, the lower limb axis deviates medially, and in valgus deformation, the lower limb axis deviates laterally. Created with BioRender.com, adapted from Bellezza (2020).*

According to Bellemans et al. (2012), the normal HKAA can be from 0 to 3 degrees of varus. Varus and valgus alignment increase medial and lateral load, respectively (Sharma et al., 2001). By causing uneven load distribution across the knee joint, deformities in the alignment of the lower limb are a proposed risk factor for knee pathologies, including accelerated progression of knee osteoarthritis (Brouwer et al., 2007). (Ramazanian et al., 2022)

2.1.2 Osteoarthritis

Osteoarthritis (OA) is a degenerative disease of the entire joint, including bone, cartilage, ligaments, fat, and synovium (Arden & Nevitt, 2006; Jang et al., 2021). Osteoarthritis most commonly affects the knee, hip, hand, and spine joints. In an osteoarthritic knee joint, the protective cartilage in the joint degrades more rapidly than new cartilage is formed. Also, structural changes appear in the subchondral bone, and the joint is inflamed (Heidari, 2011). Gradually, these processes lead to changes that can be seen in plain radiographs: knee joint space narrowing, changes in knee alignment, and presence of osteophytes (Heidari, 2011). (Jang et al., 2021)

Advancing age is a risk factor for all OA types is age, but the risk factors especially for knee OA are female sex, obesity, previous knee injury, and deformities of the lower limb (Heidari, 2011). According to Sharma et al. (2001), varus and valgus alignments can each be both a cause and a result of progressive knee OA. Varus or valgus alignment that predates knee OA may be due to genetic, developmental, or posttraumatic factors. Knee alignment that results from knee OA is usually due to loss of cartilage and bone height (Sharma et al., 2001). In both situations, uneven load on the knee joint causes uneven cartilage stress and degradation, worsening OA in the knee. This creates a self-feeding cycle that progressively worsens knee OA. (Sharma et al., 2001)

The main symptoms of OA are pain, morning stiffness, reduced range of motion, and functional limitations (Heidari, 2011; Jang et al., 2021). Osteoarthritis is diagnosed based on the symptoms and clinical examination. Imaging is not needed to establish a clinical diagnosis of OA, but OA structural changes have traditionally been assessed with radiographs (Hunter & Bierma-Zeinstra, 2019). A widely used classification system for the diagnosis of knee osteoarthritis based on plain radiographs is the Kellgren and Lawrence (KL) classification system (Jang et al., 2021). KL classification is based on the above-described changes in the structure of the joint: Joint space narrowing and formation of osteophytes (Kellgren & Lawrence, 1957). KL classification divides the severity of OA into five categories presented in Table 2.1.

Table 2.1. Kellgren-Lawrence classification system (Kellgren & Lawrence, 1957).

Grade	Description
0	Definite absence of X-ray changes of osteoarthritis
1	Doubtful joint space narrowing and possible osteophytic lipping
2	Definite osteophytes and possible joint space narrowing
3	Moderate multiple osteophytes, definite narrowing of joint space, some sclerosis, and possible deformity of bone ends
4	Large osteophytes, marked narrowing of joint space, severe sclerosis, and definite deformity of bone ends

Osteoarthritis is deemed present at grade 2, although of minimal severity (Kellgren & Lawrence, 1957). While the Kellgren-Lawrence grading is popular, it may not be ideal for

assessing individual knees due to overlapping values, complicating treatment decisions (Felson et al., 2011). Therefore, it is better for orthopedic surgeons to base treatment decisions on the physical measurements of each knee. This approach may lead to more accurate diagnoses and better treatment outcomes for patients (Kohn et al., 2016).

In the treatment of osteoarthritis, the approaches include increasing physical activity, weight loss for those who are overweight, and pain medication. The goal of medication for osteoarthritis is to relieve pain and maintain functionality, as there are no medications that repair cartilage damage. In severe cases when joint pain disrupts normal movement and daily activities, or if there is pain at rest that is not alleviated by medication, knee OA is treated with total knee arthroplasty (Hsu & Siwiec, 2023). (Jang et al., 2021)

2.1.3 Total knee arthroplasty

Total knee arthroplasty (TKA), also called total knee replacement, is a surgical procedure to resurface a damaged knee (Hsu & Siwiec, 2023; Sales 2018). The goals of knee joint replacement surgery are (1) to alleviate chronic and severe knee pain, (2) to restore the normal knee alignment, and (3) to restore joint function, and eventually improve quality of life by resurfacing the parts of the knee joint that have been damaged and correcting possible deformities (Manjunath et al., 2015). Total knee arthroplasty includes removing the damaged cartilage and bone, and implanting three new components: the patella component, tibia component, and femur component (Figure 2.3). A polyethylene spacer is inserted between the components to create a smooth, gliding surface (Foran & Manner, 2021; Hsu & Siwiec, 2023). Knee anatomy and knee prosthesis are presented in Figure 2.3.

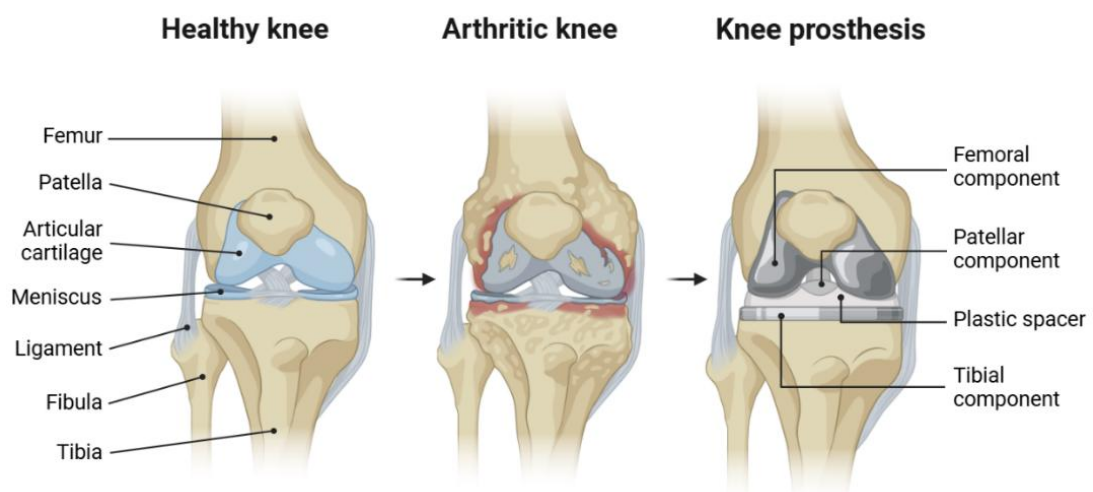


Figure 2.3. Total knee arthroplasty. Healthy knee anatomy before TKA; Bone, cartilage, and ligament wear due to knee osteoarthritis; Knee prosthesis after TKA. Created with BioRender.com according to Foran & Manner (2021).

Successful TKA requires thorough preoperative assessment of lower limb alignment, bone quality, ligament stability, and range of motion. Measuring the HKAA is particularly crucial for planning the surgery, since achieving proper alignment ensures correct prosthesis positioning, restoring normal limb mechanics, and reducing complications (Jang et al., 2021; Manjunath et al., 2015). Accurate implant alignment, together with anatomical balance, provides more successful clinical outcomes and a longer duration of the prosthetic material. According to Manjunath et al. (2015), even minor mispositioning of the implant can lead to early loosening, increased polyethylene wear, poor function, and eventually, increased joint stress, resulting in revision TKA.

To plan a TKA, a preoperative image exam that evidences all reference points (centers of femur head, knee joint, and tibial platform) is necessary to estimate the mechanical axis of the femur and tibia. Conventional radiographs of the lower limb are the most frequently used imaging method for planning and postoperative follow-up of TKA. (Zorzi & Miranda, 2018)

2.1.4 Radiographic imaging

Radiographs, commonly known as X-rays, utilize ionizing radiation to produce 2D images of the internal body structures (Hirvasniemi et al., 2014). X-ray imaging was developed based on the discovery made by Wilhelm Röntgen in 1895 (Röntgen, 1896). X-rays are created when electrons are accelerated from a heated cathode toward a dense metal anode by applying a high voltage. The heated cathode releases electrons through a process called thermionic emission. These electrons are then driven toward the anode, where their sudden deceleration produces X-rays, a form of high-energy radiation. This radiation travels through the object being examined, and a two-dimensional detector array captures the intensity of the X-rays that pass through, allowing for image formation. (Mikla et al., 2012)

As x-ray photon passes through an object, the intensity of radiation is attenuated according to the Beer-Lambert law (equation 2.1):

$$I_x = I_0 \cdot e^{-\mu x}, \quad (2.1)$$

in which I_x is the intensity of an X-ray beam after passing through thickness x of the substance, I_0 is the initial intensity, μ is the substance's attenuation coefficient, and x is the thickness of the substance (Hirvasniemi et al., 2014; Mayerhöfer et al., 2020). Coefficient μ depends on the elemental composition of the substance and is larger for electron-dense materials like bone and smaller for electron-lucent substances like muscle, fat, and cartilage (Chen et al., 2012). In denser materials, the X-ray photons can

be completely absorbed by transferring all their energy to the inner-shell electrons of the atoms. This is known as the photoelectric effect, and it contributes to the high contrast seen in X-ray images between bone, which appears white in the film, and soft tissue, which appears black (Chen et al., 2012).

The most used imaging method for clinical assessment of knee alignment and for planning and following the success of TKA is radiographs (Marques Luís & Varatojo, 2021). Radiographic knee alignment is evaluated using a full-length weight-bearing radiograph of the lower limb with both knees in full extension, the patella facing forwards, and the X-ray beam parallel to the joint line (Babazadeh et al., 2013; Zorzi & Miranda, 2018). This is called a long-leg radiograph (LLR) (Figure 2.4).



Figure 2.4. Examples of long-leg radiographs from three different patients. Long leg radiographs can be taken from both legs or from one leg. Image quality, contrast, and positioning (standing/lying down) can vary. Overlapping bones, existing implants, and text in the radiographic images bring challenges to automatic analysis.

Plain radiographs are a standard imaging method for evaluating knee alignment because radiographs can capture the entire leg from hip to ankle in a single image, which is essential for accurately measuring the mechanical axis of the lower limb (Marques Luís & Varatojo, 2021). Since radiographs are performed while the patient is standing, it allows for weight-bearing images that reveal the knee's alignment under load (Marques Luís & Varatojo, 2021). This is crucial for understanding alignment in conditions such as varus or valgus deformities, as weight-bearing can affect joint positioning and reveal misalignment that may not appear when lying down (Marques Luís & Varatojo, 2021). Radiographs are also fast, affordable, and widely available, making them a practical choice for routine alignment assessments and follow-up imaging (Hirvasniemi et al., 2014).

Despite all the good aspects, radiographs provide only a 2D representation of the leg, making it challenging to visualize complex anatomical structures or rotational deformities accurately. Overlapping bones can obscure details or distort alignment assessment. Also, variations in patient positioning during imaging, image quality, and measurement inconsistencies can affect diagnostic accuracy. (Sheehy & Cooke, 2015; Zorzi & Miranda, 2018)

2.1.5 Coxa Hospital for Joint Replacement

Coxa Hospital for Joint Replacement (Coxa) is the only hospital in Finland focusing on joint replacement surgery. Coxa is a public hospital owned by five wellbeing services counties (main owners: Pirkanmaa, Päijät-Häme, and Kanta-Häme wellbeing services counties). Treatment at Coxa can be accessed with a doctor's referral from anywhere in Finland, under the free choice of treatment place (*vapaan hoitopaikan valinta*) under the Health Care Act. In 2023 there were 15,613 knee replacement surgeries performed in Finland, of which Coxa performed 4,101 (3,914 primary joint replacements, 187 revision joint replacements). (THL, 2024).

A key part of TKA planning and monitoring TKA outcomes is determining knee alignment, which is currently performed semi-automatically from plain LLRs using computer-aided diagnosis (CAD) tools. Orthopedists mark specific points on a radiograph (center of femoral head, center of knee joint, and center of tibial plafond) to let the program measure alignment automatically based on the surgeon's markings. However, this approach presents some challenges: different orthopedists may place these points in slightly different locations, causing variability in the results. Even the same surgeon might produce different alignment measurements on different days. Additionally, manual alignment marking is time-intensive, and with rising surgery volumes, reducing the workload for surgeons would be beneficial. (Przystalski et al., 2023)

Coxa prioritizes scientific research as a core part of its operations. The research at Coxa primarily focuses on two areas: (1) joint replacement surgeries and their outcomes and factors, and (2) medical and healthcare perspectives in joint surgery and patient care (CoxaPro, 2024). Due to growing demand for joint replacements and the variability and time-intensivity of manual alignment determination, Coxa aims to enhance its use of imaging data, especially plain radiographs. Medical image processing algorithms and AI models are in development to improve diagnostic accuracy and facilitate imaging data processing for research. This Master's thesis represents an initial step towards achieving these goals.

2.2 Medical image processing

Medical imaging (MI) is the procedure used to attain images of the interior of the body for medical uses to identify or study diseases. Techniques to enhance, analyze, and extract relevant features from images are called image processing. Medical image processing (MIP) aims to extract diagnostic information from medical images, such as radiographs, magnetic resonance images (MRIs), or computer tomographies (CTs), by using computers. (Abdallah et al., 2019)

During the last decades, the paradigms of medical image processing have changed. Medical image processing began with manual interpretation of analog X-ray films in the 1960s, rapidly followed by the advent of digital imaging like CT and MRI (Bercovich & Javitt, 2018). Digital image acquisition enabled basic computer-aided processing such as filtering and segmentation. In the late 20th century, model-based and machine learning methods emerged, which allowed for more advanced analysis (Galić et al., 2023). Since the 2010s, deep learning (DL) has revolutionized the field, enabling end-to-end image analysis with high accuracy and efficiency (Mahony et al., 2020).

Key goals for MIP are image enhancement, segmentation, registration, classification, detection and quantification, and visualization (Obuchowicz et al., 2024). Image enhancement aims to improve the visual quality of an image for better analysis. Techniques used are, for example, noise reduction, contrast, and edge enhancement. These are used to make faint features more visible and prepare images for further analysis. In segmentation, an image is divided into meaningful regions like organs, tissues, or lesions using thresholding, active contours, or watershed algorithms. Image can be segmented based on intensity values (thresholding), for example, when separating bone from soft tissue in an X-ray image (Bhanage & J., 2016). In image registration, multiple images of the same subject taken at different times, modalities, or viewpoints, are aligned. This is a commonly used technique to monitor tumor growth over time. Classification aims to automatically assign labels to image regions, for example, to healthy and diseased tissue. Detection and quantification, like object or pattern recognition, are used to automatically locate and measure abnormalities or anatomical features, like HKAA, from plain LLRs. Image visualization aims to convert medical data into intuitive formats, like 3D visualization of organs, for diagnosis, planning, and communication. A more detailed chart about MIP is presented in Figure 2.5. (Obuchowicz et al., 2024)

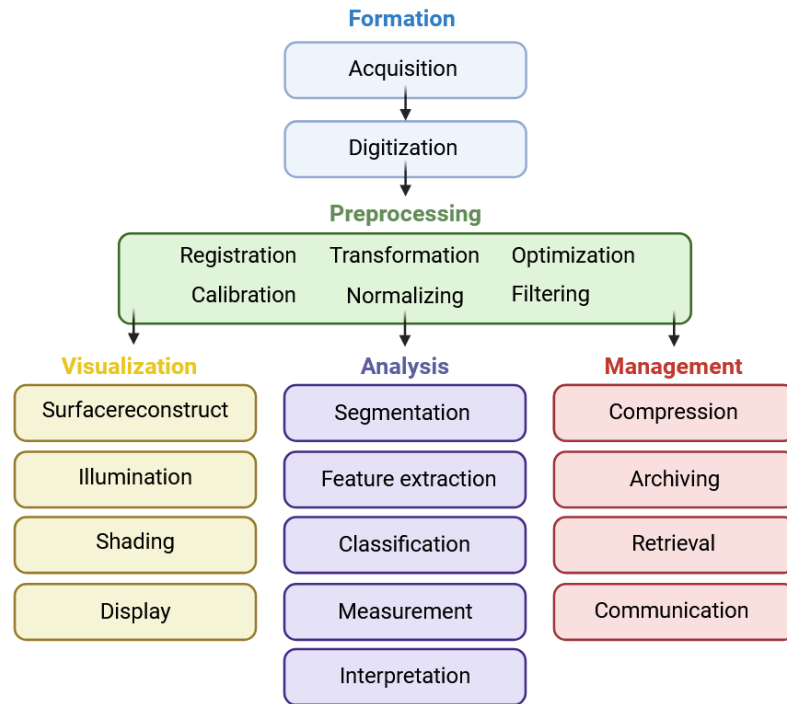


Figure 2.5. Scheme of medical image processing. Created with BioRender.com according to Deserno (2009).

With the help of MIP, clinicians can detect diseases earlier, reduce variability in human interpretation, and provide personalized treatment. Automated image analysis helps speed up workflows, allowing clinicians to focus on complex cases, and extends healthcare access through remote imaging capabilities. Image processing is essential in numerous medical fields, enhancing diagnosis, treatment planning, and research. It is used, for example, in radiology to enhance X-rays, CT, and MRI scans, allowing for clearer images of internal structures, improving diagnostics and surgical planning, and in orthopedics to process bone and joint images to assess fractures, monitor healing, and detect joint degeneration, aiding in surgical planning and treatment. (Obuchowicz et al., 2024)

However, these advancements come with challenges. Privacy and data security are the main priorities, as medical imaging involves sensitive information (Dhar et al., 2023). The use of AI also brings challenges. AI models face issues with generalizability, as they may not perform equally well across diverse populations and healthcare settings (Dhar et al., 2023). The high costs of developing and implementing AI systems can limit their accessibility, and regulatory approval processes are often lengthy and complex (Deheyab et al., 2022). Artificial intelligence methods used for medical image processing are usually computer vision-based. The following chapters will discuss the objectives, variety, and importance of computer vision in MIP and explain how automation can be applied to radiographic analysis.

2.2.1 Computer vision in medical image processing

The capacity of machines to analyze digital images by transforming them into data or extracting data from them is known as computer vision (CV) (Javaid et al., 2024). Computer vision is a subfield of artificial intelligence. Its main goal is to simulate human vision in computers, enabling them to recognize and interpret objects in images in a way similar to human perception. It involves the development of algorithms that can process, analyze, and understand images or videos. However, it outperforms human eyesight in speed and accuracy (Javaid et al., 2024). In medical imaging, CV is used to extract clinically relevant information from medical images, such as detecting anomalies, segmenting anatomical structures, and classifying diseases (Elyan et al., 2022).

Computer vision can be divided into two main fields: traditional computer vision and deep learning (DL). DL methods use neural networks, especially convolutional neural networks (CNNs), to automatically learn features from data. The key difference between traditional CV methods and DL methods is that traditional methods use hand-crafted features, whereas DL methods can learn features from the images. (Elyan et al., 2022)

Convolutional neural networks are a specific type of DL intended for image analysis. They learn to recognize patterns, like edges, shapes, or objects, in the image. CNNs can take medical images, such as plain radiographs, as input. CNNs consist of convolution layers and pooling layers. Convolution layers apply operations to the input, such as filters, to detect features in the image. These filters slide over the image and highlight patterns like edges or textures. The convolutional operation combines the input and filters, creating a feature map. The feature map shows where certain features appear and how strong they are. Between every convolutional layer, a pooling layer is applied. (Elyan et al., 2022)

Pooling layers reduce the spatial dimensions of the feature maps while retaining the most important information, thereby improving computational efficiency and focusing on high-level structures. By stacking multiple convolutional and pooling layers, a CNN can learn to detect complex features and patterns from images. After several layers of filtering and pooling, the output is flattened into a one-dimensional vector and passed into one or more fully connected layers, which make the final prediction. In radiographic analysis, the output could be a classification such as “no angular deformities detected” or “valgus deformity detected”. (Elyan et al., 2022).

Deep learning has significantly advanced the field of computer vision. CNNs learn hierarchical features automatically from raw image data, without the need for handcrafted feature design. CNNs are robust to noise and distortion in the input data, which makes

them highly effective and accurate in real-life image recognition and classification. However, certain challenges persist. Deep learning requires large, annotated datasets, which are not always possible or practical to obtain. CNNs also require significant computational resources, including high-end graphics processing units, to train and deploy. (Dhar et al., 2023)

Traditional CV relies on hand-crafted features. Algorithms are manually designed and programmed to seek specific features and relevant patterns from an image. Common techniques include edge detection, thresholding, texture analysis, and feature descriptors. Compared to DL, traditional CV is fast on simple tasks and does not need a huge dataset to perform accurate analysis. In many cases, traditional CV techniques can provide more computationally efficient solutions than deep learning approaches. (Mahony et al., 2020)

LLRs are acquired under standardized protocols (Wirtanen, 2024) with consistent orientation, standing position, full weight-bearing, and minimal variation, which reduces variability and simplifies image processing. For this reason, a traditional CV is applied in this thesis. The key anatomical landmarks, the femoral head, the knee joint, and the ankle joint, are usually well visible. The bones in LLRs (femur and tibia) follow relatively linear paths, making geometric methods useful for estimating mechanical axes and measuring the HKAA. In addition to that, a traditional CV doesn't need a large training dataset. In contrast, training deep learning models to localize anatomical landmarks would require many annotated LLRs, which are often unavailable or expensive to create.

The success of a traditional CV heavily relies on the choice and design of features. Traditional CV also struggles with variations in image modality, noise, and patient demographics. Problems with traditional CV methods might occur if the LLRs do not follow a consistent orientation, have unusual anatomy, or if the image quality is poor. Typically, these cases might include existing hip prosthesis, severe deformities, or difficulties achieving standardized image acquisition. Results will show whether traditional methods are accurate enough or should more advanced DL methods should be used in the future.

2.2.2 Automation of radiographic analysis

Radiographic imaging as a medical imaging method is one of the most accessible, popular, and widely used for the identification of a wide range of different conditions, and thus, processing radiographic images should be developed in a more automatic direction. AI has the potential to improve diagnostic accuracy, efficiency, and speed in radiography. Deep learning and machine learning have shown promising results when it

comes to image segmentation, detecting abnormalities, and making quantitative analysis. (Hardy & Harvey, 2020)

Automation in radiographic image analysis is transforming clinical workflows by reducing radiologists' workload, enhancing measurement accuracy, and improving reproducibility. In orthopedics, tasks such as determining limb alignment or calculating the hip-knee-ankle angle often rely on manual identification of anatomical landmarks, which are time-consuming and susceptible to inter- and intra-observer variability, especially in complex or low-quality images. Automated systems aim to standardize this process. According to (Ghuwalewala et al., 2021), the level of automation in medical imaging can be categorized from 0 to 5 (Table 2.2).

Table 2.2. Six levels of automation in radiologic image processing. (Ghuwalewala et al., 2021)

Level of automation	Description
0: No automation	Radiologists perform all tasks manually, no AI.
1: Radiologist assistance	Machines assist radiologists, image enhancement tools. Radiologists make all decisions.
2: Partial automation	AI assists in identifying pathologies. Final decision made by radiologists.
3: Conditional automation	AI makes some decisions without supervision under a narrow set of predefined conditions.
4: High automation	AI makes most decisions without supervision. AI request human intervention in complex cases.
5: Full automation	AI capable of end-to-end analysis of a case, from the initial diagnosis to automatic report generation.

Traditional computer vision approaches typically align with level 1 or 2, assisting with landmark detection but requiring human validation. More advanced systems using deep learning, particularly convolutional neural networks (CNNs), are pushing toward levels 3 and 4. These models can learn to localize anatomical features, detect deformities, and even suggest clinical measurements with minimal supervision, given enough high-quality training data (Nguyen et al., 2020; Wang et al., 2023).

2.3 Techniques to determine hip-knee-ankle angle

Several techniques are available to determine the hip-knee-ankle angle from plain radiographs. In the past, the angle was determined manually, but today, semi-automated and AI-assisted programs have been developed for this purpose. These will be reviewed next.

2.3.1 Manual and semi-automated methods

Hip-knee-ankle angle can be determined manually from radiographs by clinicians. Manual methods refer to situations in which clinicians manually place the anatomical landmarks (the center of femoral head, the center of knee joint, and the center of tibial) on medical images, draw lines between these points, and lastly, manually, for example, using the goniometer, measures the angle (Huang et al., 2020).

Semi-automatic HKAA determination refers to a process where both a computer algorithm and a human share responsibility in identifying anatomical landmarks and calculating the HKAA on a full-leg radiograph. In semi-automatic methods, orthopedists manually place the anatomical landmarks on medical images. These landmarks include the center of the femoral head, the center of the knee joint, and the center of the tibial plafond. Once the physician manually sets the landmarks, the software takes over. It automatically draws lines connecting these points: a line from the center of the femoral head to the center of the knee, and another from the center of the knee to the center of the ankle. The software then calculates the angle between these two lines, known as the HKA angle, and provides the measurement automatically. Examples of such programs are MediCAD and Orthopilot. (Przystalski et al., 2023)

However, the manual placement of the landmarks can be time-consuming and may vary between physicians, introducing inter-observer variability. This variation can affect the reliability of the measurement, as different physicians may interpret and place the points slightly differently, potentially impacting the accuracy of the final HKAA calculation. (Jang et al., 2021)

2.3.2 Computer vision-based methods

In addition to the traditional determination of HKAA, recent advances in AI have enabled the use of artificial intelligence for determining knee alignment from plain radiographs. Three primary deep learning approaches have emerged: landmark detection models (Jo et al., 2023; Pei et al., 2020; Tanner et al., 2024; Yan et al., 2022), segmentation-based models (Schock et al., 2020), and end-to-end angle prediction models (Kim et al., 2024;

Nguyen et al., 2020; Tack et al., 2021; Wang et al., 2023). These approaches are reviewed next.

Landmark detection models focus on identifying specific anatomical points, such as the centers of the femoral head, knee joint, and ankle, within the radiographic image. Once these landmarks are localized, geometric computations derive clinically relevant angles like the HKAA or FTA. These models typically use convolutional neural networks (CNNs) trained on manually annotated landmarks. The output is the precise coordinates of landmarks, from which angles are calculated in a post-processing step. One of their main advantages is their interpretability: because the predicted angles are calculated from visible anatomical features, clinicians can easily trace and verify the logic behind a prediction, making these models clinically intuitive and modifiable. However, their performance is highly dependent on the quality and consistency of landmark annotations used during training. If landmarks are obscured due to pathology, low image quality, or anatomical variation, the model's ability to generalize may be compromised. (Jo et al., 2023; Pei et al., 2020; Tanner et al., 2024; Yan et al., 2022)

Segmentation-based models employ CNN architectures (commonly U-Net variants) to delineate bone contours such as the femur and tibia. These segmented masks allow the identification of anatomical axes and landmarks indirectly, from which alignment angles are calculated. Segmentation thus acts as an intermediate representation between raw images and angle measurement. Segmentation-based models could be utilized for multiple tasks, like measuring knee joint space width, which is another crucial measurement to detect knee OA. The downside is that they require extensive annotated segmentation data and are more computationally demanding. Segmentation errors also directly impact final angle accuracy. (Schock et al., 2020)

End-to-end angle prediction models bypass intermediate steps and directly predict alignment angles from raw radiographs using CNN regression networks. Such models often utilize architectures like ResNet to map input images to continuous angle values or categorical deformity stages without explicit landmark or segmentation supervision. They are particularly attractive in large-scale or time-sensitive clinical settings. Nevertheless, their major drawback is reduced interpretability. Since predictions are made without explicit anatomical references, it is harder to validate the correctness of the model's decision or understand failure modes. Moreover, such models may struggle with generalization, especially when presented with anatomical abnormalities that were underrepresented or absent in the training data. (Kim et al., 2024; Nguyen et al., 2020; Tack et al., 2021; Wang et al., 2023)

AI development for knee alignment relies on large, well-annotated datasets. Public datasets like the Osteoarthritis Initiative (Branch, 2017) provide thousands of knee radiographs with associated clinical measurements, enabling broad accessibility and benchmarking (Tack et al., 2021; Wang et al., 2023). Private datasets are typically collected within hospitals or research institutions, often including long-leg radiographs (LLRs) for full limb alignment analysis (Pei et al., 2020). These datasets can range from less than 400 images (Kim et al., 2024; Pei et al., 2020; Schock et al., 2020; Simon et al., 2022), to a few thousand images (Tack et al., 2021; Tanner et al., 2024; Wang et al., 2023; Yan et al., 2022) and even over ten thousand images (Jo et al., 2023; Nguyen et al., 2020). Dataset size directly impacts model performance; larger datasets improve generalizability and robustness. However, obtaining extensive, high-quality annotations is expensive and time-consuming, especially for segmentation and landmark detection tasks.

Evaluation of AI models for knee alignment measurement generally focuses on accuracy, reliability, and clinical applicability. Mean absolute error (MAE) and mean squared error (MSE) quantify how close predicted angles are to manual measurements (Wang et al., 2023). The intraclass correlation coefficient (ICC) assesses agreement between AI and human measurements, with values near one indicating strong reliability (Jo et al., 2023). For segmentation models, the Sørensen-Dice coefficient measures the overlap between predicted and manual bone masks (Schock et al., 2020), while landmark detection models often use mean average precision to evaluate point localization accuracy (Tanner et al., 2024). Bland-Altman analysis is also used to identify systematic biases (Kim et al., 2024).

Time efficiency comparisons are also reported to demonstrate potential workflow improvements (Kim et al., 2024; Schock et al., 2020). Selecting appropriate metrics depending on the model type ensures a comprehensive evaluation of performance, supporting both technical accuracy and clinical usefulness. The choice of evaluation metrics typically depends on the model type; spatial accuracy metrics like Dice and mAP are critical for segmentation and landmark detection, while regression accuracy measures such as MAE and ICC are more relevant for end-to-end prediction models. Incorporating multiple complementary metrics helps ensure that AI solutions are both accurate and reliable enough for clinical use.

While AI models have demonstrated high levels of accuracy in HKA angle measurement, several challenges remain. High-quality annotated datasets, especially for segmentation and landmark detection, are limited and costly to produce, restricting model generalizability to diverse patient populations and imaging protocols. Anatomical variations, de-

formities, and implants can hinder model accuracy, especially for landmark-based methods requiring precise localization (Pei et al., 2020; Tanner et al., 2024). End-to-end models, while efficient, risk reduced interpretability, making clinical validation and acceptance more difficult. Furthermore, standardization of evaluation metrics and benchmarking across datasets is needed to enable fair comparisons and regulatory approval. Addressing these challenges will require larger multi-center datasets, robust algorithms resilient to real-world variability, and transparent validation to ensure safe integration into clinical workflows. (Dhar et al., 2023)

3. MATERIALS AND METHODS

3.1 Dataset

The study utilizes data from Miikka Mänttäre's thesis (Mänttäre, 2023). Mänttäre's research data includes all primary knee replacement surgeries performed at Coxa Hospital for Joint Replacement in 2018. In total, there were 2,143 primary knee replacement patients at Coxa in 2018 (Mänttäre, 2023). These patients were referred to Coxa either by primary healthcare services within the Pirkanmaa Hospital District or under the free choice of care program from other healthcare regions (Mänttäre, 2023). Before referral, a long-leg radiograph was ideally taken in a standing position. If this was not possible, the X-ray was primarily taken in a supine position. An overview of the data is presented in Table 3.1.

Table 3.1. Description of the dataset (Mänttäre, 2023).

Sex	Women	1,328 (62%)
	Men	815 (38%)
BMI (body mass index)	Lowest	17.0
	Highest	60.0
	Average	30.65
Age	Lowest	33.8 years
	Highest	93.2 years
	Average	67.3 years
Lower limb alignment	Varus	1659, max 25.5 degrees
	Valgus	442, max 24.0 degrees
	Missing	42
Kellgren-Lawrence grade	3	419
	4	1707

The dataset includes patients with tibiofemoral osteoarthritis graded at levels 3 or 4 on the Kellgren-Lawrence scale, as they made up 99.2% of the cases (2,126 patients) (Mänttari, 2023). Data on the mechanical axis of the lower limb, which is routinely measured as part of standard care and recorded in the joint replacement follow-up system, was retrieved from Coxa's data repository (Mänttari, 2023). An Excel document with mechanical axis information for each patient was created. The HUS Academic secure environment is used to interpret the data.

3.2 Data processing

Medical imaging data must be handled according to The Act on the Secondary Use of Health and Social Data (The Secondary Use Act) (In Finnish "Toisiolaki" 2019/522 2019/522). This Act came into effect in April 2019. It is designed to enable the efficient and secure processing of personal data collected in social and healthcare services, as well as for purposes such as administrative oversight, research, statistics, and official guidance within the social and health sectors. This legislation establishes standardized conditions for the secure and efficient secondary use of health and social data, ensuring the protection of individuals' privacy (Toisiolaki 2019/522).

The Secondary Use Act defines two types of data use: primary use and secondary use. Primary use of health and social data refers to the purposes for which the data was originally recorded in patient registers. This may include, for example, planning, delivering, and monitoring patients' treatment or rehabilitation. Secondary use of social and health data means using patient and register data generated in social and healthcare services for purposes other than the original reason for which the data was collected. The Secondary Use Act defines the legally permitted secondary purposes as: scientific research, statistics, development and innovation activities, official supervision and guidance, planning and investigative tasks by authorities, education, and knowledge-based management. (Toisiolaki, 2019/522)

The Secondary Use Act defines how social and healthcare data can be disclosed and used for research purposes. The Finnish Social and Health Data Permit Authority, Findata, operates as the licensing authority under the Secondary Use Act. It is responsible for granting data permits and making related decisions (Toisiolaki, 2019/522 § 44). Findata grants permit the secondary use of social and health data when the data is needed from multiple public data controllers, from the private sector, or the Kanta services (Findata, 2025) In addition to that, Findata is responsible for providing guidance, maintaining the secure data systems defined in the Secondary Use Act, and managing the secure compilation, preprocessing, and delivery of data in accordance with granted

permits (Findata, 2025). As of May 1st, 2022, individual-level social and healthcare data may only be disclosed to audited, secure environments, where the data can also be analyzed (Toisiolaki, 2019/522 § 60). This ensures the confidentiality and security of sensitive personal information. The secure environment used in this work is HUS Academic.

HUS Academic is a secure, scalable, virtual environment that complies with the Secondary Use Act. It is accessed via a web browser, making it virtually accessible from anywhere in the world as needed. The platform enables global research and is available to both the public and private sectors. HUS Academic is a Software as a Service solution. Research data, in accordance with the research group's permit, is delivered to HUS Academic, where the research team can then conduct their analysis within the secure environment. (Turunen, 2024)

The radiographic images are uploaded to the secure environment. Radiographic images are transferred as DICOM (Digital Imaging and Communications in Medicine) format because it is an international standard to transmit, store, retrieve, print, process, and display medical imaging information. To ease the image processing, DICOM format is converted to the JPEG format. Each image is sequentially coded with a unique, six-numbered, patient-specific identifier number, for example, 123456. Data on the mechanical axis is linked to each image and recorded in a separate Excel file. This data is then transferred to HUS Academic, where the analysis is conducted.

3.3 Manual measurement procedure and narrowing the dataset

Orthopedist's measurements were performed during 2018 semi-automatically. Orthopedists placed joint locations for each LLR, and software provided angle measurements based on those coordinates. Measurements were collected into an Excel file, containing only information about the patient number (a six-digit code to anonymously individualize each patient) and corresponding HKAA value for each patient. From the dataset (Table 3.1), approximately 800 alignment images were transferred to the secure environment due to the long delivery time of the extensive dataset. For some patients, there was more than 1 HKAA value, presumably from the right and left knee. Since they weren't labeled and could not be tracked, which value is from which leg, these patients were left out of the analysis. Due to the same reason, patients with two-legged radiographs but only one HKAA value in the Excel file were left out, resulting in a dataset with 363 patients with one-legged LLR and one unambiguous HKAA value in the Excel file.

3.4 Image preprocessing

Images need to be preprocessed in order to enhance the image quality and improve the visibility of bone and joint structures. Preprocessing is executed for both analyzed images and templates. Used preprocessing methods used are noise reduction, histogram equalization, and edge detection, which are now explained in more detail. Python 3.12.0 was used for image preprocessing, template matching, angular deformity detection, and HKAA measurement.

3.4.1 Noise reduction

Noise in medical images refers to random variations in pixel intensity, which obscure or distort the true image content. Noise in medical imaging is often caused by various sources, such as electronic sensor interference, low light, limitations in the imaging system, or random movement during imaging (Manson et al., 2019). Especially in radiographic imaging, the main source of noise is secondary radiation, which arises from scattered X-rays from the machine and the object before reaching the film (Manson et al., 2019).

It is essential to filter the noise in the radiographic images because it potentially increases the false detection rate in computer vision-based methods. Since all computer vision methods are easily affected by the noise in the radiographic image, it is essential to filter out the noise to prevent false detection caused by it. Noise reduction techniques like smoothing filters are applied to clean the image, preserving as much detail as possible while eliminating unwanted noise. These filters can reduce contrast among neighboring pixels and thereby reduce the amount of noise in the image (Manson et al., 2019). Gaussian blur is one method to reduce noise in the image, making the image smoother. Gaussian blur smooths an image by averaging each pixel's value with its neighbors, using weights defined by a Gaussian function (equation 3.1) (Sharda, 2021). The Gaussian function for 2-dimensional images is determined as follows:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \quad (3.1)$$

in which x is a horizontal distance from the center pixel, y is the vertical distance from the center pixel, and σ is the standard deviation that controls the blur (Sharda, 2021).

A kernel is a small matrix (for example, 3×3 or 5×5 pixels) built using the equation 3.1 above. In a Gaussian blur, the kernel slides over the image, and in each position, the weights from the kernel are multiplied by the corresponding pixel values in the image. These weighted values are summed, and the result replaces the center pixel. The center

of the kernel always has the highest weight, and nearby pixels influence the result more than distant ones. This process is repeated for every pixel in the image, causing the image to become smoother by reducing noise and fine detail. Larger kernels create a stronger blur because they include more distant pixels in the calculation. This is demonstrated in Figure 3.1. (Gonzalez & Woods, 2018)



Figure 3.1. *Gaussian blur. Original image on the left (A). Gaussian blur with kernels 5 (B), 9 (C), and 13 (D).*

It can be seen from Figure 3.1 that a moderate rate (kernel 5) of Gaussian blur makes bones more visible from the background and fades the background noise (Figure 3.1B). Excessive blur (Figure 3.1C and D) makes it difficult to distinguish bones, especially the femoral head. Gaussian blur must be used carefully to ensure that distracting noise is removed without losing important details in the image.

3.4.2 Histogram equalization

A histogram is a graphical representation of the distribution of pixel intensities in an image (Leong, 2023). Radiographic images are grayscale images, so the histogram is only for intensity values, not for color channels. Pixel intensity can vary from 0 (black) to 255

(white) (Leong, 2023). Histogram equalization can be used if images are represented by a narrow range of intensity values or if pixels are clustered in the middle of the available intensity range (Gonzalez & Woods, 2018). Through histogram equalization, the intensities can be distributed to the full range of intensities evenly to enhance details and ensure overall balance in the image (Gonzalez & Woods, 2018). The effect of histogram equalization can be seen in Figure 3.1.

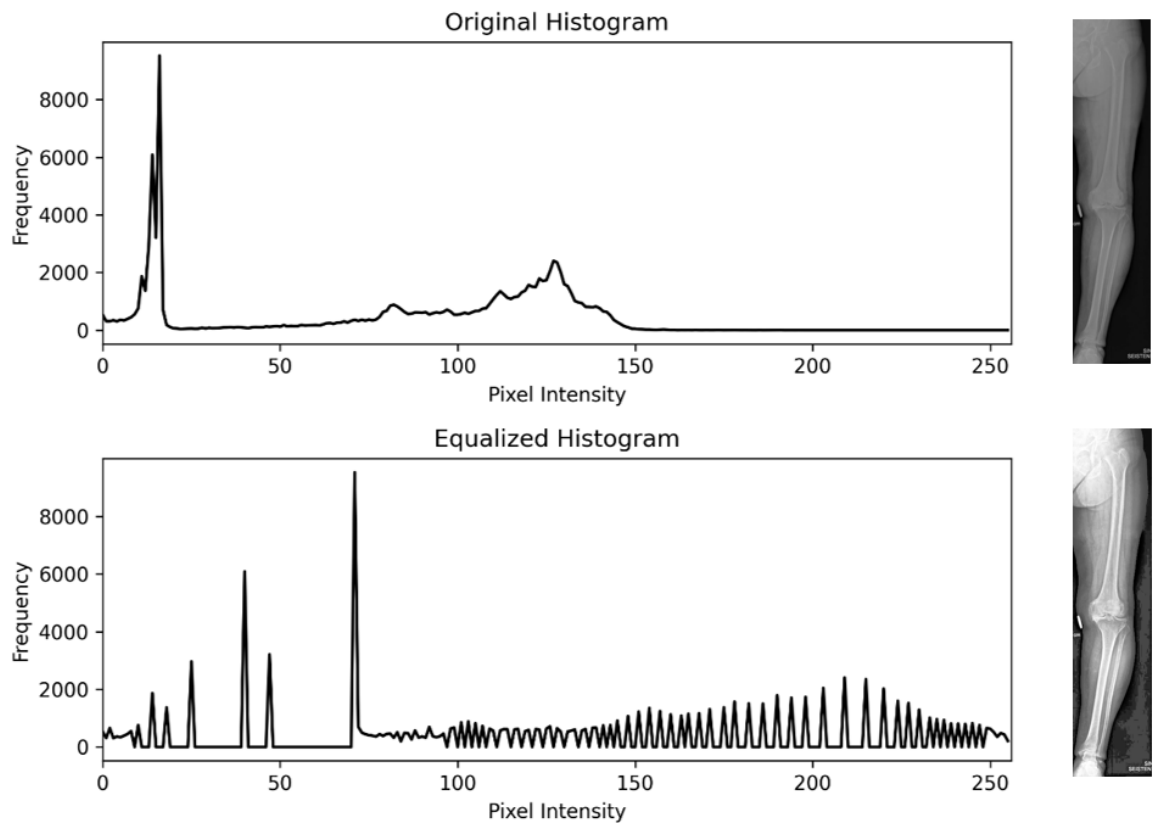


Figure 3.2. Histogram equalization spreads the pixel intensity values to a wider range in the histogram plot, improving the contrast in an image. Histogram equalization makes bone structures and joint locations more visible, and they stand out better from the background. This makes analysis easier and more reliable.

The standard deviation of the pixel intensities is measured to determine if an image needs histogram equalization or not. Histogram equalization is applied if the standard deviation of pixel intensity is lower than or equal to 50, meaning that there's not much variation or spread in the brightness values of pixels. Histogram equalization is not used for images with a standard deviation of pixel intensity larger than 50, since in those images the bone and joint areas are distinct enough to obtain reliable results.

3.4.3 Edge detection

Canny edge detection was developed by John F. Canny in 1986 (Canny, 1986), and it is a popular edge detection algorithm based on the intensity gradient of the image. After noise reduction, the image is filtered with a Sobel kernel in both horizontal and vertical directions to receive the edge gradient (equation 3.2) and direction for each pixel (equation 3.3) as follows:

$$\text{Edge Gradient } (G) = \sqrt{G_x^2 + G_y^2}, \quad (3.2)$$

$$\text{Angle}(\theta) = \tan^{-1}\left(\frac{G_y}{G_x}\right). \quad (3.3)$$

Edge gradient G consists of horizontal gradient G_x and vertical gradient G_y , and describes how strong the edge is at each point. Angle θ indicates the direction of the edge. These indicate how quickly pixel values change in the image. Large gradients indicate sharp changes, and these changes are possible edges. (Gonzalez & Woods, 2018)

An important stage is to decide whether an edge is really an edge or not. For this, two threshold values, minimum and maximum, are needed. Any edge with edge gradient more than the maximum value is sure to be an edge, and those below the minimum value are sure to be non-edges. Those who lie between these two thresholds are classified as edges or non-edges based on their connectivity. If they are connected to "sure edge" pixels, they are considered to be part of the edges. Otherwise, they are discarded and the image is left with strong edges. Canny edge detection is visualized, and the impact of setting different minimum and maximum threshold values is shown in Figure 3.3. (Gonzalez & Woods, 2018)

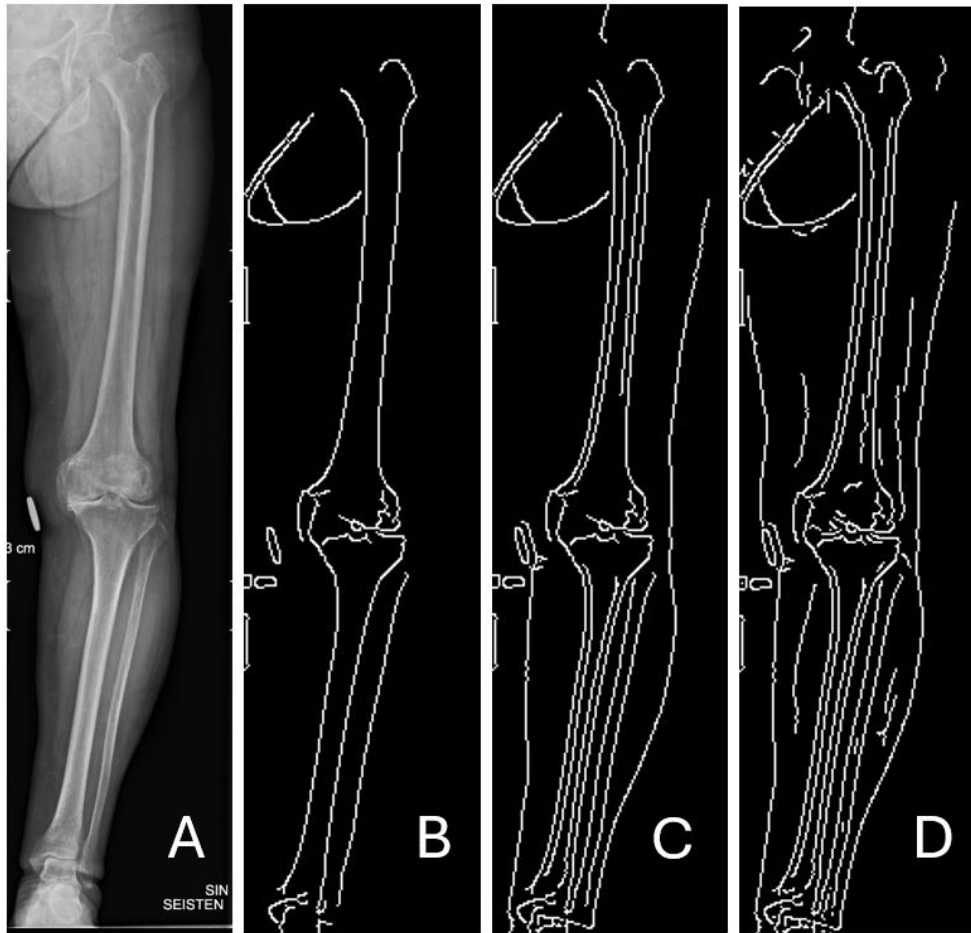


Figure 3.3 Canny edge detection and the impact of threshold values. The original image is on the left (A). The minimum threshold value stays the same (50), but the maximum decreases from left to right: 200 (B), 150 (C), 80 (D).

It can be seen from Figure 3.3 that the lower the maximum threshold value is, the more edges are detected. The placement of the femoral head partially inside the pelvic bone challenges edge detection. It can be seen from Figure 3.3D that the femoral head is detected at low threshold values but at the same time, the unwanted extra edges in the ankle area increase. Finding the correct minimum and maximum threshold values is critically important for highlighting necessary areas without negatively affecting other areas.

3.5 Template matching

Template matching is a computer vision method dealing with visual content. It compares a smaller image, a template, against larger image regions. It slides the template over the image and calculates how well it matches each position. The best matching area is determined as the position of the target (Gonzalez & Woods, 2018). The Squared Difference Normalized (Python: `cv.TM_SQDIFF_NORMED`) method is used in this work, and it is conducted according to equation 3.4:

$$R(x, y) = \frac{\sum_{x', y'} (T(x', y') - I(x + x', y + y'))^2}{\sqrt{\sum_{x', y'} T(x', y')^2 \cdot \sum_{x', y'} I(x + x', y + y')^2}}, \quad (3.4)$$

in which I denotes the pixel in the image under the template at position (x, y) , T denotes the pixel value in the template, and R denotes the result score at position (x, y) (OpenCV, n.d.). For each position, template matching places the template over the image, computes the difference between each pixel in the template and corresponding image region, squares the differences and sums them up. The result is normalized to the range from 0 to 1. The smallest value means less difference between template and image, indicating the best match (Gonzalez & Woods, 2018). Figure 3.4 demonstrates a template matching with three templates for one image.

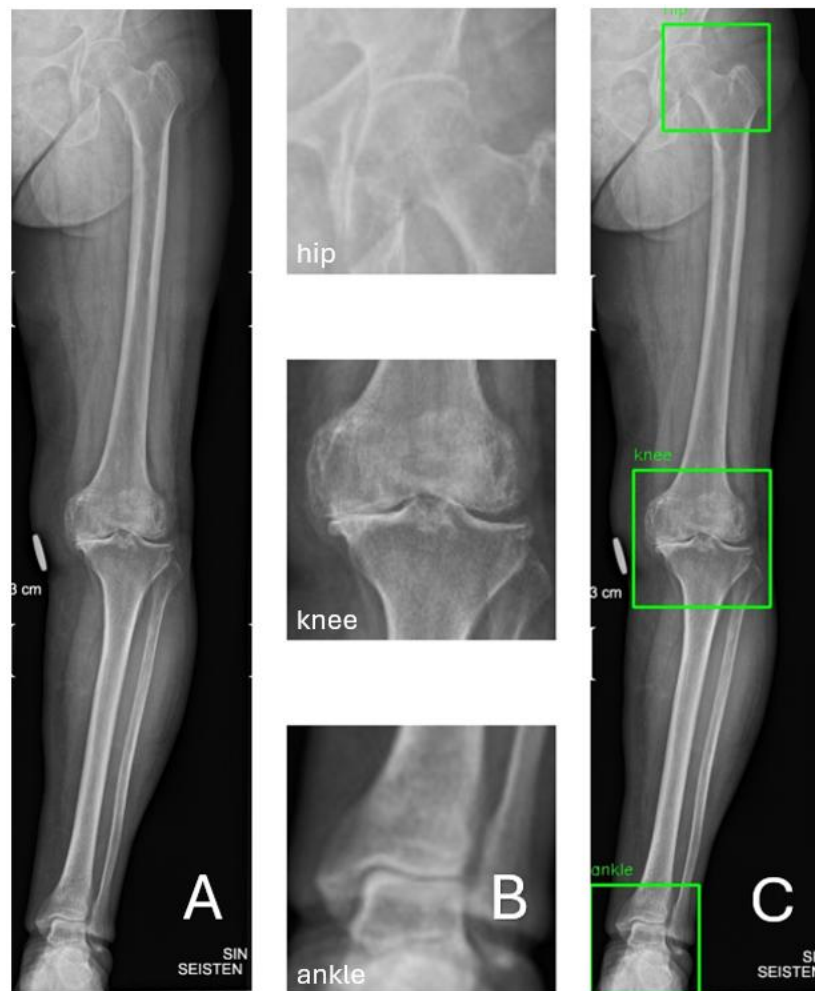


Figure 3.4. Template matching. A is the original image, B are the template images of the hip, knee, and ankle, and C is the matched template sites in the original image.

Figure 3.4 presents used template matching method used for this work. Template matching is used to locate three joint locations: the hip, knee, and ankle. It is important to keep

in mind that osteoarthritic knees, ankles, and hips can look very different and have different positions and angles. In addition to that, the quality of the radiographic images can also vary. Also, the anatomy of the right and left legs differs from each other. Therefore, one template for each joint type is not enough to detect the variety of knees, hips, and ankles. Thus, atlas-based template matching is used.

An atlas is a standardized reference set of images representing typical anatomy, often derived from multiple subjects to provide a generalizable template for each matched location (hip, knee, and ankle). Normal individual variation in human lower limb structure presents a significant challenge to the atlas selection. Building an atlas to represent standardized lower limb structure, covering variations in limb structure as well as possible, is a time-consuming process. This study tries to optimize the use of the atlas for knee, hip, and ankle detection.

An important factor when building an atlas is trying to include a variety of samples in the atlas to cover the diversity of lower limb anatomical structures. These include, for example, templates from right and left limbs, templates from different quality radiographs, varus and valgus deformities, existing knee or hip implants, and taking possible sex-related structure variations into account. Six atlases are built: For the right and left hip, knee, and ankle. At first, from 5 to 10 good-quality, clear-structured images were chosen for each atlas, and joint areas were cropped to correspond to the correct location. Atlases were supplemented with several templates during the process to meet the needs of the dataset. The number of templates in each atlas gives an idea of how easy or difficult each joint is to detect. The fewer templates needed, the easier the joint is to find, and the more templates needed, the more difficult the joint is to find or generalize to fit a particular set.

To increase reliability and decrease false matches, and to speed up the template matching process, and according to human anatomy, the search area for template matching is restricted. The image is split vertically into three regions: the top, middle, and bottom regions. Hip must be matched in the top region, knee in the middle region, and ankle in the bottom region. This prevents matches in unwanted, likely false regions. The program uses the middle points of each matched template and draws lines between the points (from hip to knee and from knee to ankle) and calculates the HKAA using the laws of cosine (equation 3.5).

$$HKAA = \arccos\left(\frac{v_1 \cdot v_2}{|v_1||v_2|}\right). \quad (3.5)$$

In equation 3.5, the HKAA is calculated as the angle between vectors. Vector v_1 represents the direction from the femoral head to the knee joint, and v_2 represents the direction from the knee joint to the ankle joint. Dividing the dot product of these two vectors ($v_1 \cdot v_2$) by the product of the magnitudes of the two vectors and finally taking the inverse cosine (arccosine) of the results gives the HKAA in radians. The result is converted to degrees for clinical interpretation. This angle is referred to as the predicted value in further analysis, and it is compared to the observed value determined by an orthopedist.

HKAA values measured by an orthopedist are either positive or negative, indicating either varus or valgus deformities, respectively. Before determining deformity type, the algorithm needs to know which leg (left or right) is being analyzed in the radiographic image. Since images contain only one limb and it is not marked in metadata or elsewhere whether the limb is left or right, the script infers this based on template matching scores. Templates are separated for each side (right and left), and for each joint algorithm matches both left-side templates and right-side templates and saves the best match score for both sides. The side with the higher total score is assumed to be the correct leg.

Deformity type is determined in two ways, referred to as methods A and B. Method A is to compare the matched knee location to the mechanical axis (MA) of the leg (Figure 3.5). In radiographs of the left leg, if the knee coordinate is right to the MA the deformity is varus and the angle is marked as positive, and if the knee coordinate is left to the MA, the deformity is valgus and the angle value is marked as negative. Same with right legs: If the knee coordinate is right to the MA, deformity is valgus, and if the knee coordinate is left to the MA, deformity is varus. This method is directly based on the medical determination of varus and valgus positions.

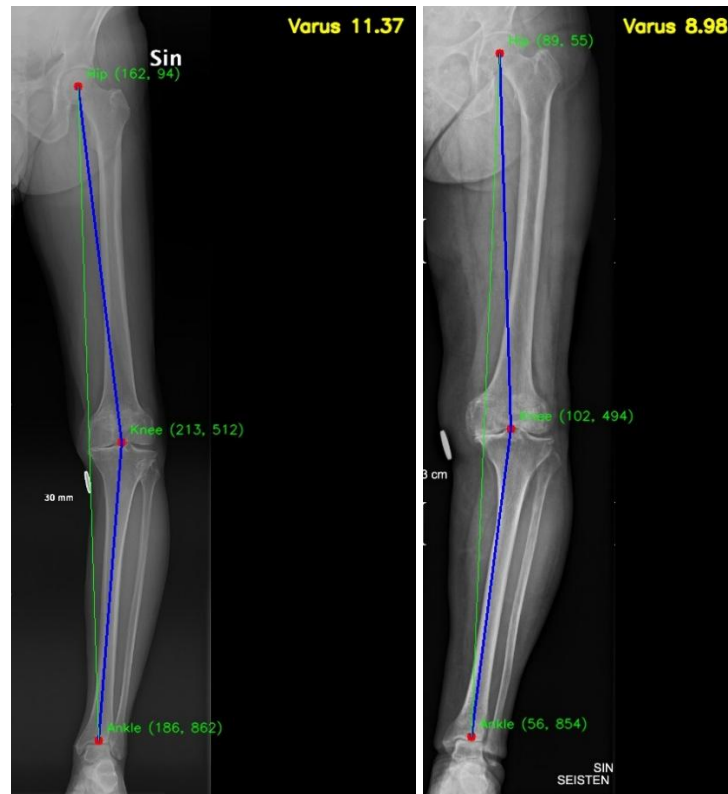


Figure 3.5. Detection of varus and valgus deformities, method A. The location of the matched knee coordinate relative to the mechanical axis (green line) of the lower limb determines deformity type. Here, both radiographs are from the left leg, and the matched knee coordinate is located right of the MA. This results in varus deformity and a positive HKA value.

Method B is to compare matched knee x-coordinates to matched ankle x-coordinates. For left legs, if the matched knee x-coordinate is larger than the ankle x-coordinate, the deformity type is varus, and if the knee x-coordinate is smaller than the ankle x-coordinate, the deformity type is valgus. For the right leg, valgus deformity is when the knee x-coordinate is larger than the ankle coordinate, and varus is when the knee x-coordinate is smaller than the ankle coordinate. In Figure 3.5 can be seen that in both left leg images, the knee x-coordinate is larger than the ankle coordinate, indicating a varus position.

This analysis was conducted for 363 long-leg radiographs, and results were saved to an Excel file next to the orthopedist's measurements. If both are either positive (varus) or negative (valgus), the algorithm has detected the right form of deformity. If the HKA value is close to 0 degrees, errors in deformity detection increase. Only those HKA values for which the deformity is correctly determined are used in the statistical analysis of the model. This is because the wrong sign incorrectly increases the error between the measured angle values. For 363 long-leg radiographic images of 330 MB, the analysis lasted 102 minutes. The analysis speed is approximately four images per minute.

3.6 Statistical analysis

The goodness and reliability of the model are assessed using various statistical methods: mean average error (MAE), root mean square error (RMSE), intraclass correlation coefficient (ICC), Pearson coefficient (PCC), and the coefficient of determination (R^2).

3.6.1 Error

Root mean square error (RMSE) and mean average error (MAE) are used to compare predicted and observed values, and measure how much error there is between two datasets. RMSE is defined as follows (equation 3.6):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}}, \quad (3.6)$$

in which P_i represents the predicted values (algorithm measurements), O_i represent the observed values (orthopedist's measurements), and n is the total number of observations (Chugh, 2024). RMSE values can vary from 0 to positive infinity. Low RMSE values indicate that the model fits the data well and has more precise predictions close to observed values. Conversely, higher values suggest more error and less precise predictions, with poor predictive accuracy. (Chugh, 2024)

Mean absolute error (MAE) is a measure of errors between paired observations. MAE is defined as follows (equation 3.7):

$$MAE = \frac{\sum_{i=1}^n |O_i - P_i|}{n}, \quad (3.7)$$

and can be described as an arithmetic average of the absolute errors. Like RMSE, low values of MAE indicate less error and high values more error. (Chugh, 2024)

According to Coxa's orthopedists, the measurement difference between two orthopedic surgeons can be up to three degrees. In this study, that three-degree margin is used as the threshold for excellent results, as it would mean the algorithm is as accurate as human experts. A deviation twice that size (3-6 degrees) is considered the limit for average performance, while larger errors are classified as poor results. RMSE values below 6 degrees and MAE values below 3 degrees are ideal. Errors exceeding 3 degrees may begin to negatively impact surgical planning.

Although both RMSE and MAE measure errors between two datasets and share the same unit (degrees), they differ in how they treat those errors. RMSE squares the differences, which means it gives more weight to larger errors, making it particularly sensitive to outliers. In contrast, MAE treats all errors equally, providing a straightforward average

that is less influenced by extreme values. These differences are important when interpreting model performance, which is why both metrics are used in this study.

3.6.2 Correlation

Three methods for measuring the correlation between orthopedists' and algorithms' measurements are used. These are the Pearson correlation coefficient (PCC), the intra-class correlation coefficient (ICC), and the coefficient of determination (R^2). PCC measures linear correlation, and it is a number between -1 and 1 that measures the strength and direction of the relationship between two variables (Berman, 2016). The closer the PCC value is to 0, the smaller the correlation is. The PCC is determined as follows in equation 3.8:

$$PCC = \frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O}) \sum_{i=1}^n (P_i - \bar{P})}}, \quad (3.8)$$

in which in which O_i represents the observed values values (orthopedist's measurements), P_i represent the predicted values (algorithm measurements), \bar{O} and \bar{P} are respectively the mean values of the two variables, and n is the total number of observations (Berman, 2016). PCC value can be interpret as presented in Table 3.2.

Table 3.2. Interpreting the PCC value (Schober et al., 2018).

PCC value	Correlation
Less than 0.2	Poor
Between 0.2 and 0.6	Moderate
Between 0.6 and 0.8	Good
Greater than 0.8	Excellent

In this study, PCC over 0.60 is attempted.

The intraclass correlation coefficient (ICC) describes how strongly units in the same group resemble each other. Used ICCs are ICC(2,1) (Two-way random effects, absolute agreement, single rater) presented in equation 3.9 and ICC(3,1) (Two-way mixed effects, consistency, single measurement) presented in equation 3.10.

$$ICC(3,1) = \frac{MS_{subjects} - MS_{error}}{MS_{subjects} + (k - 1) * MS_{error}}, \quad (3.9)$$

$$ICC(2,1) = \frac{MS_{subjects} - MS_{error}}{MS_{subjects} + (k - 1) * MS_{error} + \frac{k}{n}(MS_{raters} - MS_{error})}, \quad (3.10)$$

In equations 3.9 and 3.10 the $MS_{subjects}$ is the mean square for subjects, the MS_{error} is the mean square for error, k is the number of measurements (2), and n is the number of paired data values (Koo & Li, 2016). ICC(2,1) describes how closely values match, called “absolute agreement”, and ICC(3,1) describes how consistently values vary together, called “consistency” (Koo & Li, 2016). With these, it is possible to measure how closely this algorithm mimics this specific experiment (ICC(3,1)) and estimate how well the method would agree with other experts if extended to a broader setting (ICC(2,1)). According to Koo & Li (2016), the value of an intraclass correlation coefficient can be interpreted as presented in Table 3.3.

Table 3.3. Interpreting the ICC value (Koo & Li, 2016).

ICC value	Correlation
Less than 0.50	Poor
Between 0.5 and 0.75	Moderate
Between 0.75 and 0.9	Good
Greater than 0.9	Excellent

In this study, an ICC over 0.8 is attempted. ICCs are determined using pingouin `intraclass_corr()` function from the pingouin statistical package (<https://pingouin-stats.org/build/html/index.html>).

The coefficient of determination (R^2) is also used to assess how effectively the independent variables in a linear regression model account for the variation in the dependent variable (Chugh, 2024). The coefficient of determination is determined as follows (equation 3.11):

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}. \quad (3.11)$$

R^2 ranges from 1 to 0. 1 means the model predicts the data perfectly, 0 means the model doesn't explain any of the variation, and values between 0 and 1 show how good the model is: the closer the value is to 1, the better the model explains the data (Chugh, 2024). In this study, R^2 over 0.7 is attempted.

4. RESULTS

This section presents the results of the study. Results are divided into 4 sections: Results related to template matching, statistical analysis, graphs, and visual presentations.

4.1 Template matching

Prior to having access to the secure environment, preprocessing methods were tested and applied for three anonymized “test” JPEG images outside the environment. After having access to the secure environment and processing images there, some differences were noticed between the image preprocessing of test images and DICOM-formatted images in the actual dataset. DICOM images have higher pixel count, making edge detection prone to errors despite a Gaussian blur. Histogram equalization produced good results in increasing contrast in critical joint areas. Thus, only histogram equalization was used for image preprocessing.

In atlas-based template matching, there are multiple templates for each matched location. These templates should cover all possible anatomical and physiological changes in the image, so that all leg alignments are taken into account. Six atlases were created: an own atlas for left and right sides for all three joints (hip, knee, and ankle). The number of templates in each category can vary depending on human anatomy, and while trying to achieve the best results. The number of templates in each category is presented in Table 4.1.

Table 4.1. Number of templates in each atlas.

Atlas	Number of templates
Ankle Right	10
Ankle Left	8
Hip Right	18
Hip Left	18
Knee Right	9
Knee Left	11

For 363 plain long-leg radiographs, 18 ankle templates, 36 hip templates, and 20 knee templates were used. Hip templates were needed twice as much compared to required ankle or knee templates. Ankle and knee templates were needed in almost the same amount. This indicates that the hip was the most difficult joint to detect since it is surrounded by overlapping bones and other tissues. Knee and ankle were easier to detect, and based on the number templates, it is not possible to say which one is easier or more difficult to detect from the two.

Time for analyzing 363 radiographs is 102 minutes, approximately 3-4 images in one minute. Analysis time is affected by the number of radiographs, the preprocessing methods used, the number of templates, and the area for template matching (1/3 of the image for each joint).

Methods A and B were used to detect the angular deformity type. Method A compared the matched knee location to the mechanical axis of the lower limb, and method B compared the matched knee location to the matched ankle location. Method A correctly identified the angular deformity in 60% of the images, and method B 80% correctly identified the angular deformity in 80% of the images. If the HKAA value is close to 0 degrees, errors in deformity detection increase. Of 363 single-leg LLRs, 100 patients (28%) have valgus deformity, the rest have varus deformity. From these, 80% (294/363) of the deformity checks made by the algorithm with method B were correct. Thus, 294 images were used in statistical analysis.

4.2 Statistical analysis

The model is statistically evaluated using mean absolute error (MAE), root mean square error (RMSE), Pearson correlation coefficient (PCC), and intraclass correlation coefficient (ICC). The algorithm's measurements are compared to the orthopedist's measurements. Statistical analysis is conducted for results that received successful angular deformity detection with method B (comparing ankle coordinate to knee coordinate). If angular detection went wrong, the patient was left out of the statistical analysis. Results are shown in Table 4.2.

Table 4.2. Comparing statistical analysis between orthopedist measurements and method B, the correct angular deformity values are obtained. Analysis is conducted for both deformity types together and for varus and valgus deformities separately, to compare algorithm performance between alignment types.

Statistical analysis method	Both deformities	Varus	Valgus
MAE	4.5 degrees	4.1 degrees	5.4 degrees
RMSE	5.6 degrees	5.1 degrees	6.7 degrees
PCC	0.85	0.23	0.62
ICC(2,1)	0.76	0.16	0.39
ICC(3,1)	0.84	0.23	0.61

The algorithm showed an MAE of 4.5° and an RMSE of 5.6°. According to Table 3.2, this indicates a moderate level of average deviation from the expert measurements. The Pearson correlation coefficient (PCC) was 0.85, reflecting a strong linear relationship between the two methods according to Table 3.3. The intraclass correlation coefficients (ICC) were 0.76 for ICC(2,1) and 0.84 for ICC(3,1), indicating good agreement between the algorithm and the orthopedist according to Table 3.4. In more detail, an ICC(3,1) being 0.84 indicates that the algorithm and rater are fairly consistent with good internal reliability, but ICC(2,1) being smaller (0.76) refers to systematic differences, for example, one always rating slightly higher or lower than the other.

4.3 Graphical analysis

To assess the accuracy and agreement between the algorithm and reference measurements, a regression analysis and Bland-Altman analysis were conducted. This approach allows evaluation of both the linear relationship and potential systematic differences between methods. The scatter plot for all results is presented in Figure 4.1, and the Bland-Altman plot in Figure 4.2. Varus and valgus alignments are also plotted separately in Figures 4.3 and 4.4, respectively.

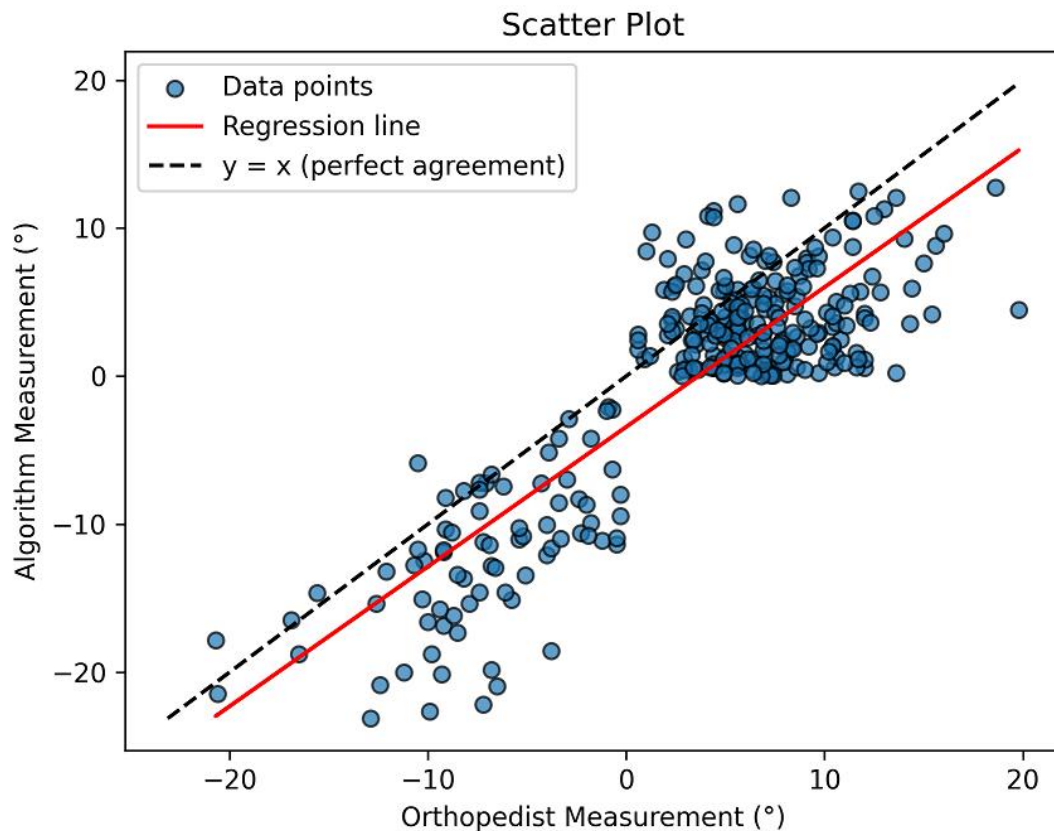


Figure 4.1. Scatter plot of the results. The slope of the regression line is 0.94, and R^2 is 0.72.

From Figure 4.1 can be seen that the data is divided into two clusters, indicating varus and valgus deformities, since the vicinity of location (0,0) is quite empty. Slope 0.94 indicates that the algorithm tends to slightly underestimate larger values, and R^2 0.72 indicates that 72% of the variability in the algorithm outputs can be explained by the orthopedist's measurement. This is considered a strong positive relationship, although not perfect, since some variance remains unexplained.

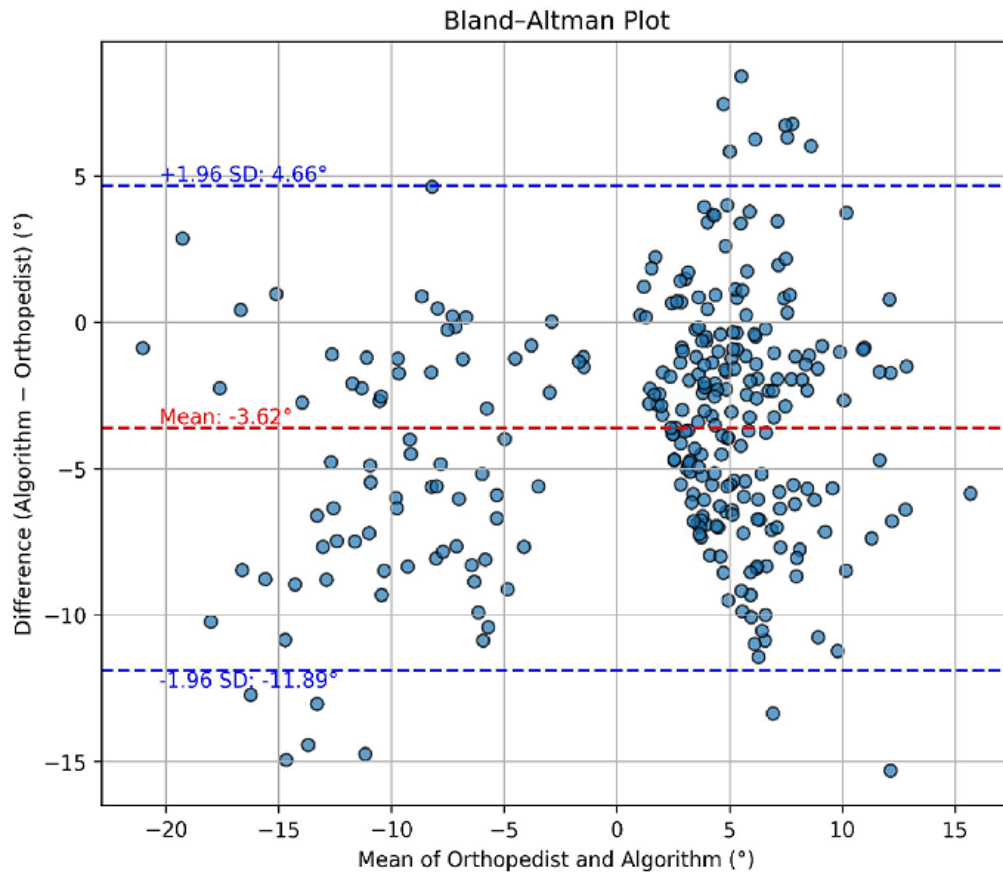


Figure 4.2. Bland-Altman plot of the results. The Bland-Altman analysis showed a mean difference of -3.6° , with limits of agreement ranging from -11.9° to 4.7° .

The mean difference in the Bland-Altman plot differs from the MAE value. This mean difference indicates that the algorithm's measurements are consistently 3.6 degrees lower than the orthopedist's measurements. The limits of agreement ranged from -11.9° to 4.66° , suggesting that 95% of individual differences between the two methods fall within this range. While the bias is relatively small, the width of the limits (approximately 16.5°) indicates high variability.

To compare the algorithm's performance in varus and valgus angle calculations, the scatter plot and Bland-Altman plot were also created separately (Figures 4.3 and 4.4).

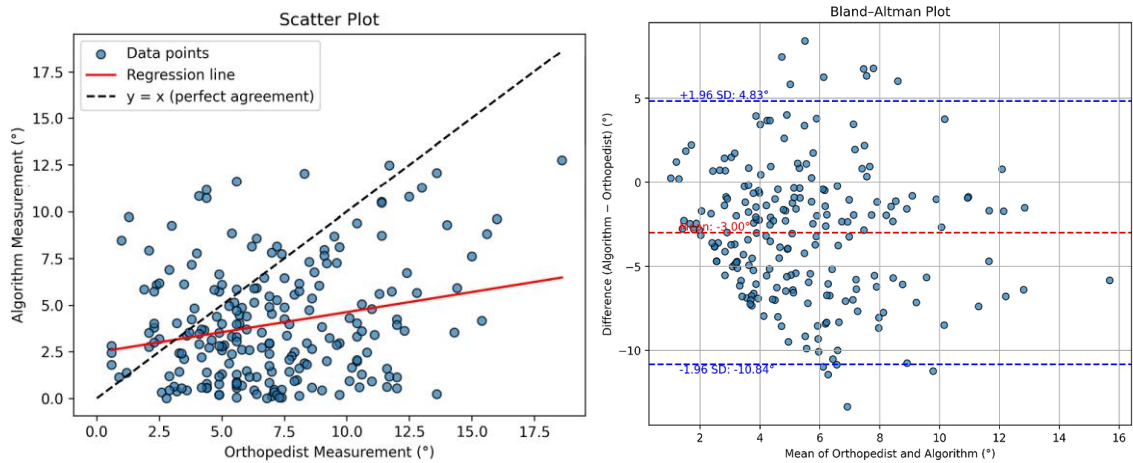


Figure 4.3. Varus analysis. The slope of the regression line is 0.22 and R^2 is 0.06. The Bland-Altman analysis for varus angles showed a mean difference of -3.1° , with limits of agreement ranging from -11.1° to 4.9° .

According to the scatter plot for varus angle values (Figure 4.3), the slope value of 0.22 and R^2 of 0.06 indicate a very weak correlation and little sensitivity to individual patient differences in varus deformities. According to the Bland-Altman plot (Figure 4.3) and similarly to Figure 4.2, varus angles are systematically underestimated. Limits of agreement have a wide range, meaning that large individual deviations can occur.

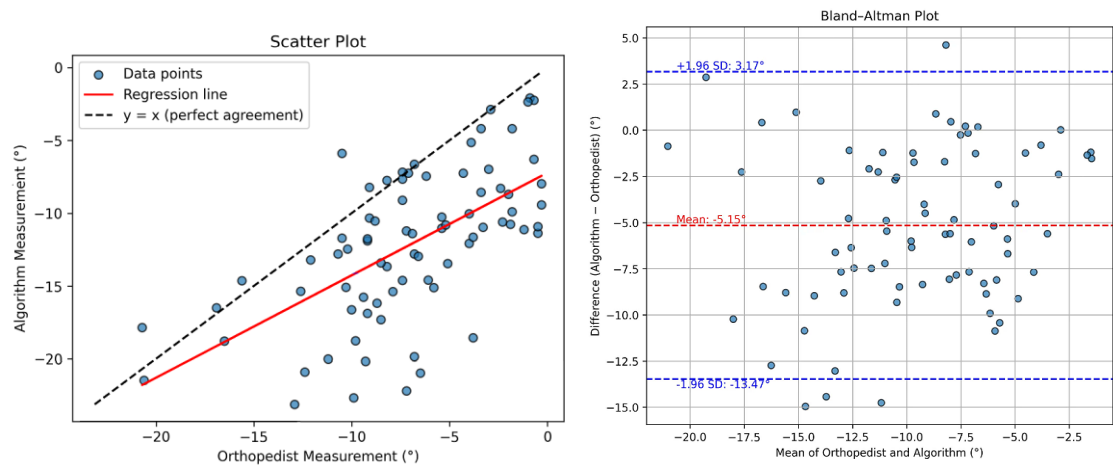


Figure 4.4. Valgus analysis. The slope of the regression line is 0.70, and R^2 is 0.38. The Bland-Altman analysis for valgus angles showed a mean difference of -5.2° , with limits of agreement ranging from -13.5° to 3.2° .

According to the scatter plot for valgus angle values (Figure 4.4), the slope value of 0.70 and R^2 of 0.38 indicate a moderate correlation. The model is partially tracking changes in patient data. According to the Bland-Altman plot for valgus angle values (Figure 4.4), the mean difference of -5.2° indicates larger systematic underestimation.

4.4 Visual review of images

Since there are no coordinates for each measurement point (middle of hip, knee, and ankle) and the analysis is conducted only based on the measured HKAA values, there are several risk factors that can potentially lead to errors in joint detection and angle calculation. It is important to get an idea of which joint is the most difficult to detect, and which features of both the image and the leg affect the success of the analysis. Thus, some analyzed radiographic images are presented next. They are grouped to present a successful analysis (Figure 4.3), difficulties in locating the femoral head (Figure 4.4), and difficulties in locating other joint areas (Figure 4.5). Visual presentation concretizes recurring challenges that come to joint location from plain, long-leg radiographs and helps to understand differences in image quality, leg positioning, and the impact of existing implants.

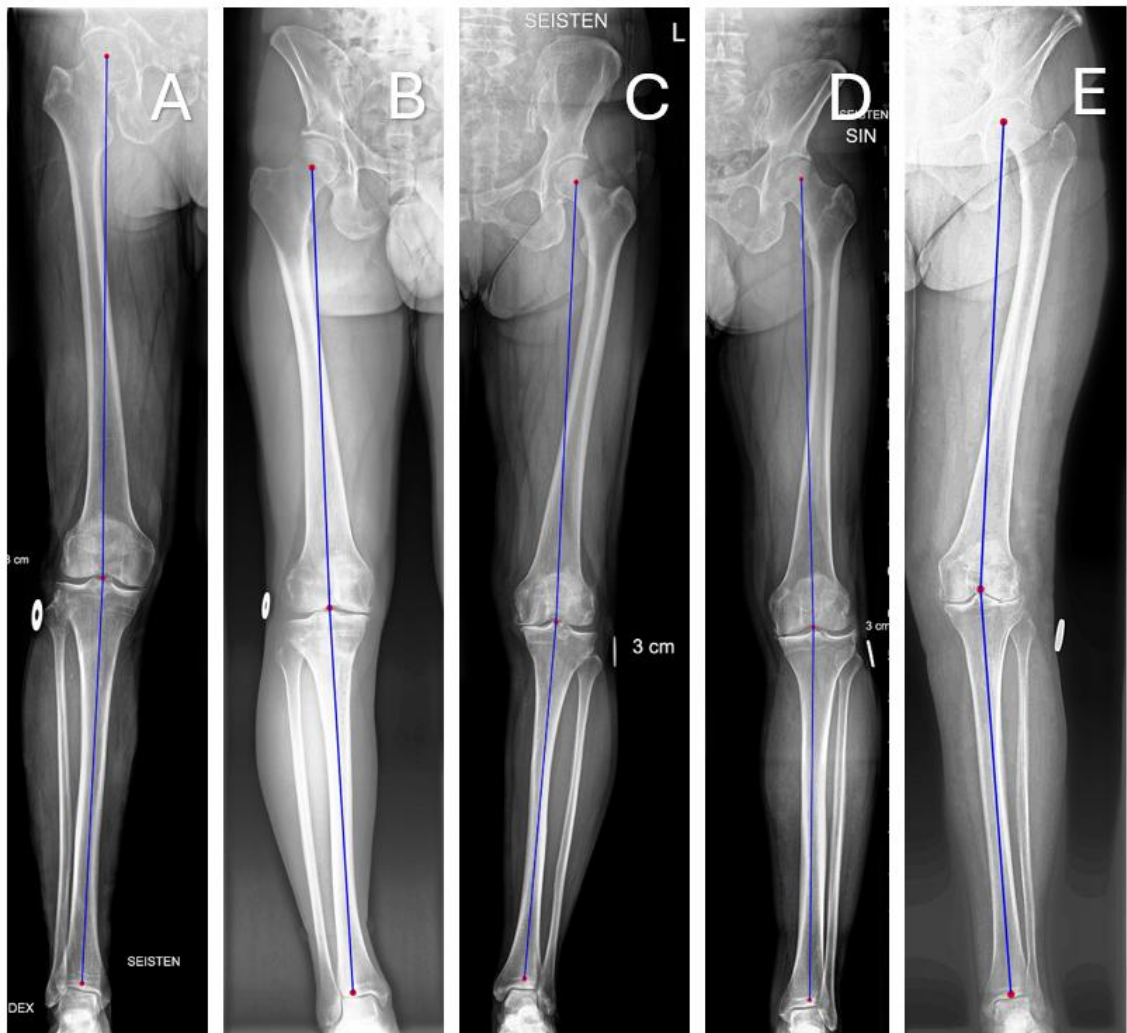


Figure 4.5. Succeed analysis. All joints are correctly located, angular deformity is correctly determined, and the HKAA value is measured correctly. Images have good quality, position, and clear bone appearance, which helps with the analysis.

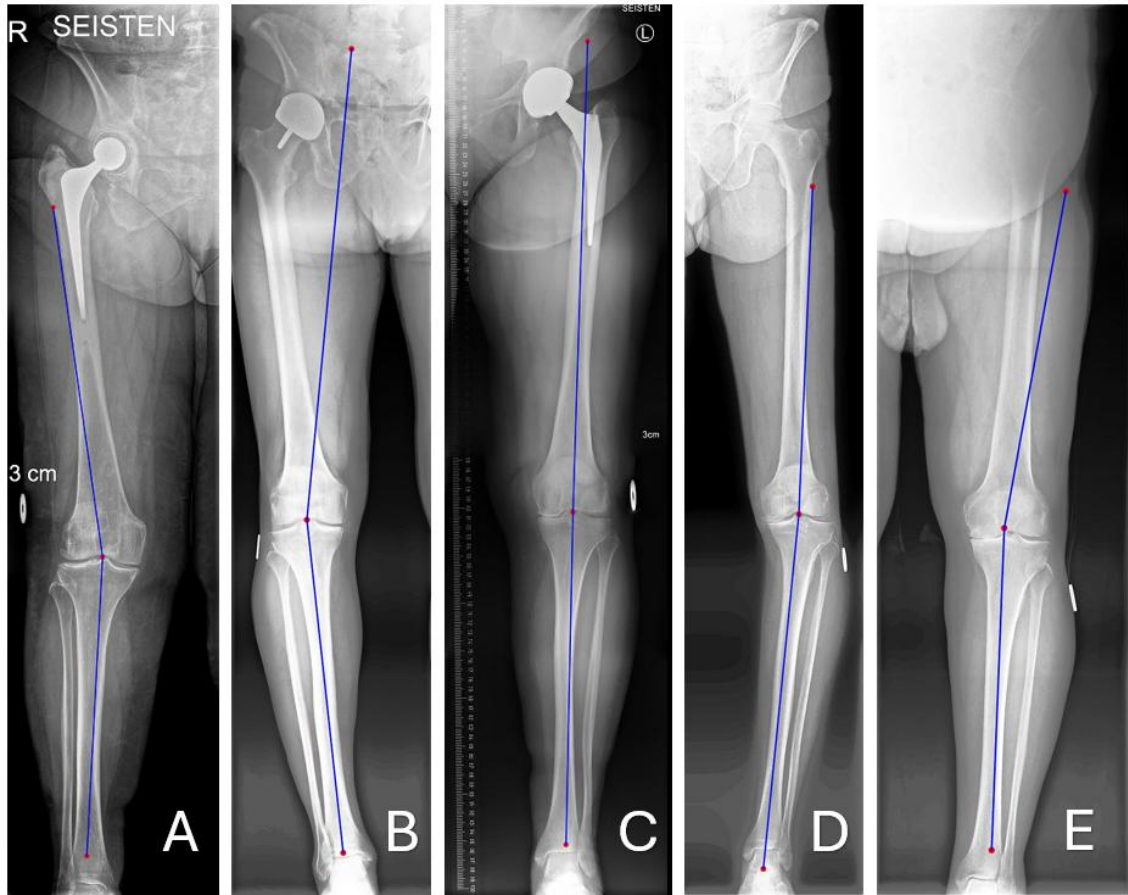


Figure 4.6 Issues with hip joint location. The hip was the most difficult joint to locate. The reason for that is its location in proximity to other bones and tissues (D and E). The most relevant spot to detect, the head of the femur, sits partly inside the pelvis. Another major factor that interfered with the detection of the hip joint location is the possible hip prosthesis (A, B, C). The sizes and shapes of hip prostheses can vary so widely that hip prostheses are difficult to generalize in template matching.

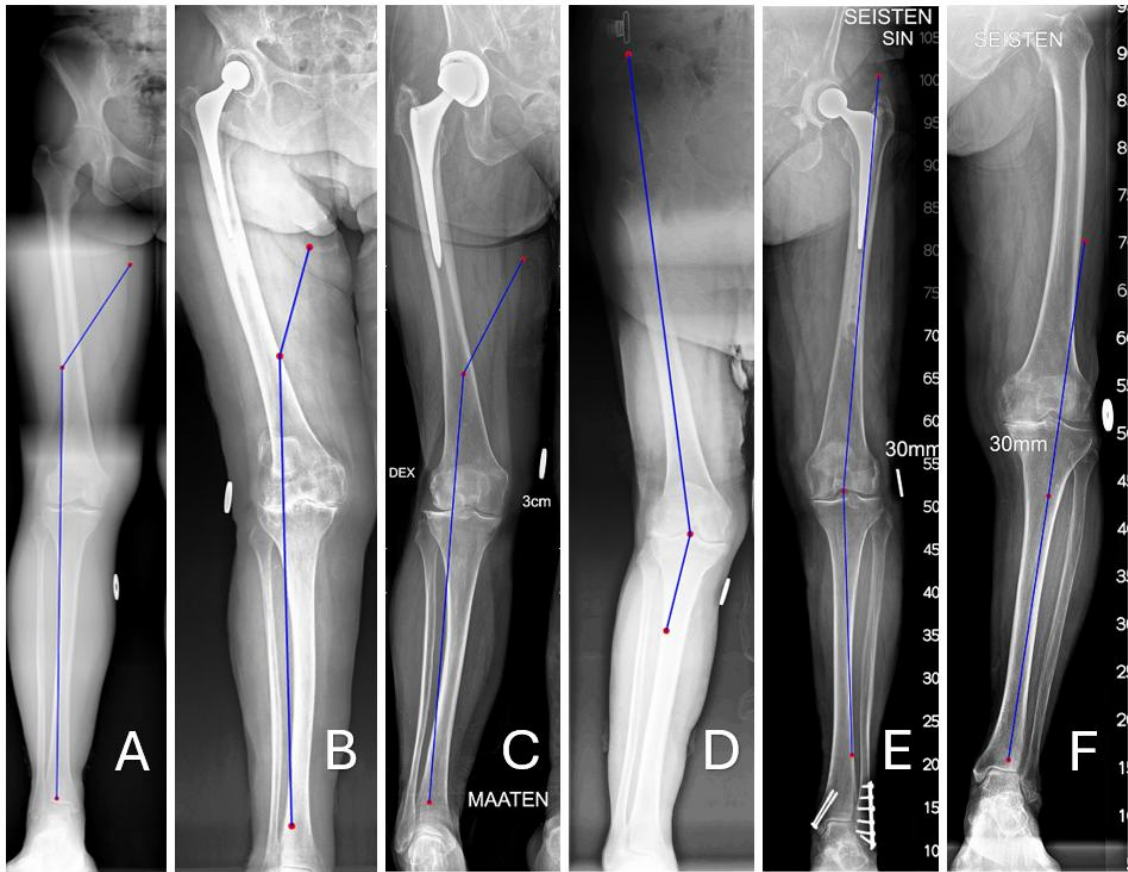


Figure 4.7. Issues with locating several joint areas. Image quality has a huge effect on detecting joint areas. If image quality is poor, bones and joint areas don't stand out as they should, and automatic analysis is not possible (A and D). If osteoarthritis has gotten worse, the joint space between the femur and tibia is small or non-existent. Because the joint space and its typical shape are essential in identifying the knee joint, their absence affects the location of the knee (B). Prostheses or screws in knees or ankles can also interfere with the analysis, making the joint site look untypical (B, C, E). Sometimes text is placed on the critical joint areas, covering them and interfering with the template matching (F). Patient position also matters. If a radiograph is taken while lying down, the position of the joint may change and not be as favorable for analysis as in an image taken while standing (C).

Based on a visual review of analyzed images showing the matched locations, other joint areas are easier to locate than others. The hip joint area is the most difficult joint to locate due to its location partly inside the pelvis, and prevalence of hip prosthesis in the older generation, since the average age for TKA, according to Table 3.1, was 67 years. Locating the ankle and knee was more successful, even though both have their difficulties. Poor image quality makes it difficult to locate all joint areas. If even one of the required locations is incorrect, it greatly affects the reliability of the angular deformity analysis and measuring HKAA.

5. DISCUSSION

The dataset consisted of radiographs from over 2,000 patients across multiple healthcare centers, introducing natural variability in image quality, contrast, and positioning. This diversity is beneficial for training a robust and generalizable algorithm. The dataset includes both men and women, a wide age and BMI range, and a good balance of left and right legs. Both varus and valgus deformities are represented, although valgus cases are fewer. While the original dataset was large, only 363 radiographs were used in the final analysis. Despite this, the sample size remains relatively same as in similar studies (Lee et al., 2024; Pei et al., 2020; Schock et al., 2020; Simon et al., 2022), although datasets of many thousands of images have also been used (Nguyen et al., 2020; Tanner et al., 2024; Wang et al., 2023; Yan et al., 2022).

The dataset consists of long-leg radiographs taken before the total knee arthroplasty, meaning that only “severely degenerated” knees (KL grades 3-4) are included. However, the Kellgren-Lawrence (KL) classification reflects only joint space narrowing and the presence of osteophytes, it does not provide information about the alignment of the lower limb. Therefore, even a knee with a high KL grade can have either a near-normal or highly abnormal deformity angle. Nevertheless, the appearance of the knee joint space in radiographs significantly affects the analysis, particularly whether there is a large gap or direct contact between the femur and tibia. For this reason, it would be beneficial to include radiographs of healthy or mildly degenerated knees (KL grades 0-2) in the dataset to improve the algorithm’s generalizability.

To evaluate the performance of the developed template-matching-based algorithm for estimating the hip-knee-ankle angle (HKAA), several statistical metrics were calculated: mean absolute error (MAE), root mean square error (RMSE), Pearson correlation coefficient (PCC), and intraclass correlation coefficients (ICC(2,1) and ICC(3,1)). In addition, a scatter plot and a Bland-Altman plot were drawn from the dispersion of the results.

The MAE of 4.5° indicates that, on average, the algorithm's HKAA predictions deviate from the orthopedist-defined values by 4.5° . While this shows that the algorithm is in reasonable proximity to expert measurements, it still leaves room for improvement, especially if clinical-grade precision is the target. Deep learning studies have received MAE values below 1.7° (Nguyen et al., 2020; Pei et al., 2020; Wang et al., 2023). The RMSE value of 5.6° reinforces this interpretation. Because RMSE gives greater weight to larger

errors, this result suggests that in some cases, the algorithm may produce significant outliers.

The Pearson correlation coefficient (PCC) of 0.85 demonstrates a strong linear relationship between the algorithm's estimates and the reference values. This suggests that, despite individual errors, the algorithm tends to follow the same directional trend as the expert measurements. The intraclass correlation coefficients (ICC) provide further insight into agreement rather than just correlation. ICC(2,1), which reflects agreement under a two-way random effects model (suitable for generalizability to other raters or datasets), was 0.76, indicating good but not excellent agreement. ICC(3,1), which assumes a fixed set of raters and assesses consistency rather than absolute agreement, was higher at 0.84, reflecting very good reliability within this specific dataset and rater pair, similar to (Simon et al., 2022). Studies utilizing deep learning have received ICC values over 0.9 (Pei et al., 2020; Tanner et al., 2024).

Based on the regression analysis and Bland-Altman plot, the algorithm tends to underestimate the HKAA by approximately 3.6° compared to measurements made by the orthopedic specialist. This consistent underestimation strongly suggests that the algorithm places the hip joint closer to the greater trochanter rather than the true center of the femoral head. As a result, the calculated angle becomes smaller and approximates the femorotibial angle, which is defined using anatomical axes rather than mechanical axes. The Bland-Altman plot also revealed a high level of variability (16.5°), which exceeds clinically acceptable limits and is much higher than in relative studies, in which variability is closer to 3° and mean below 1° (Nguyen et al., 2020; Pei et al., 2020; Schock et al., 2020). Therefore, the agreement between the algorithm and expert measurements is inadequate for clinical use. To improve the results, hip localization must be significantly refined.

When comparing the algorithm's performance on varus and valgus deformities, notable differences emerge in both accuracy and reliability. For varus, the MAE and RMSE were lower at 4.1° and 5.1° , respectively, compared to 5.4° and 6.7° for valgus, indicating greater numerical accuracy. However, the correlation metrics suggest limited agreement with reference values in the varus case: PCC was 0.23, and intraclass correlation coefficients (ICC(2,1) = 0.16 and ICC(3,1) = 0.23) were notably low. Conversely, the algorithm showed stronger consistency in the valgus analysis, with higher PCC (0.62) and ICCs (0.39 and 0.61), despite higher errors. Regression and Bland-Altman analyses supported these findings, with valgus showing better linear fit ($R^2 = 0.38$ vs. 0.06) and a slightly wider range of agreement. Overall, while the algorithm was more accurate for varus angles, it demonstrated better reliability and agreement with reference values for

valgus angles. These differences suggest that developing separate algorithms tailored to varus and valgus deformities may improve overall performance and clinical applicability.

Determining angular deformity was one of the most relevant parts of the success of the analysis. Two methods were tested: Method A, based on the mechanical axis, and method B, based on the relative x-coordinates of matched knee and ankle landmarks. When evaluated against ground truth deformity types, method A correctly classified 60% of the images, while method B achieved 80% accuracy. The better performance of method B reflects its reduced dependency on locating the femoral head. Since femoral head location was the most difficult to detect and had presumably the greatest error, it affected the definition of the mechanical axis of the lower limb. If this goes wrong, the angular deformity detection also goes wrong, which can be seen in the low success rate in method A. Although method B outperformed method A, neither method reached the level of reliability desirable for clinical use.

A major factor affecting the performance of the model is that the study compares HKAA values determined by the orthopedist and the algorithm. This metric is highly sensitive to errors, as calculating the HKAA requires correctly identifying the leg side (right/left), the type of deformity (varus/valgus), and the precise locations of all three joints: the femoral head, the knee center, and the tibial plafond. An error in any of these variables can lead to a significant deviation in the HKAA value. Ideally, each image would include orthopedist-defined coordinates for all joints, providing direct reference points for the algorithm to match. For more reliable and comparable results, a well-annotated dataset could be created in which one or more orthopedic experts mark the coordinates of the anatomical landmarks (hip, knee, ankle) once or multiple times. This would enable the analysis of intra- and interobserver variability and allow the algorithm's accuracy and error patterns to be evaluated more precisely. Such an approach would also help to identify which joint is easiest or most difficult to detect, something that HKAA comparison alone cannot fully reveal. Since only the HKAA value was available in this study, the model's performance could be assessed solely based on that angle. This inevitably leads to weaker results and potentially large deviations from the orthopedist's value, even in cases where the joint localization, such as for the ankle and knee, was highly accurate. This also increases the risk of false-positive results, where the final angle appears reasonable despite inaccurate joint detection.

Another important limitation is how the reference HKAA values were assigned. In this dataset, several patients had multiple LLRs, yet only one HKAA value was provided per patient. This assumes the deformity angle remains constant across all images, which is

not true in practice. HKAA should ideally be considered image-specific rather than patient-specific, as even small differences in patient positioning, rotation, or acquisition parameters can result in measurable changes in angle. If each radiograph had its own independently verified HKAA value, errors introduced by mismatches between image features and the assigned ground truth could be reduced. Consequently, part of the observed prediction error may stem not from model performance itself, but from the mismatch between image-specific features and patient-level annotations. Future datasets should ensure that angle measurements are aligned with individual images to enable more precise training and evaluation.

DICOM-formatted images are valuable for algorithm development, as they allow for the inclusion of metadata directly linked to each image. Typically, this metadata contains patient information such as sex, age, date of imaging, and device specifications. However, the potential of metadata could be utilized even more extensively to support image analysis. For instance, information such as the side of the limb (left/right), the type of deformity (varus/valgus), existing hip prosthesis (yes/no), and even the angle measurement itself could be stored to facilitate comparison and analysis. Based on the metadata, images could be directed to the correct image processing pipelines, each pipeline tailored to detect certain features in an image. The more precisely images can be classified before angle calculation, the more accurate atlases can be created. For example, if the metadata indicated the presence of a hip prosthesis, the algorithm could selectively use templates from a prosthesis-specific atlas to improve accuracy. Additionally, the anatomical coordinates of key joints, discussed in an earlier section, could also be included in the metadata to support training and evaluation.

As this kind of information was not available in the current dataset's metadata, the images were converted to JPEG format to simplify processing. This limited the ability to take full advantage of metadata during analysis, and image quality may have slightly degraded. Moreover, the conversion to JPEG introduced an additional processing step, which slowed down the workflow and could have been avoided by working directly with DICOM images. In general, the large number of templates, the large size of the wading area, the large number of images, and the multi-step preprocessing slow down the analysis. Therefore, it is important to create an atlas that is as compact but comprehensive as possible, and to target their use correctly, as well as to choose the preprocessing methods as efficiently as possible, avoiding unnecessary work steps in the analysis.

Traditional computer vision techniques, such as template matching used in this work, can be effective in controlled conditions with consistent image quality. In this work, they

provided a reasonable starting point for identifying anatomical landmarks when the radiographs were clear and well-aligned. However, these methods are limited in their ability to generalize. They are sensitive to variations in image contrast, patient positioning, anatomical differences, and the presence of deformities or implants. Deep learning offers a more powerful and flexible solution. Convolutional neural networks can learn complex patterns directly from large datasets and adapt to the variability present in real clinical images. They are better at handling noise, outliers, and subtle visual features, making them more suitable for tasks that require precision and robustness. With proper training and validation, they can provide results that are not only more accurate but also more consistent. Therefore, while traditional methods work as a baseline and results demonstrate a promising foundation, deep learning represents the most promising path forward for building clinically viable and scalable solutions in radiographic analysis.

6. CONCLUSIONS

This study developed an atlas-based template matching model to detect the hip, knee, and ankle joints from long-leg radiographs, determine angular deformity, and calculate the hip-knee-ankle angle. The algorithm successfully identified anatomical landmarks to determine HKAA, showing moderate agreement with orthopedist-defined values. Statistical metrics, including a mean absolute error of 4.5° and a Pearson correlation coefficient of 0.85, indicate promising but not clinically sufficient performance. Notably, the hip joint was the most challenging to localize accurately, contributing significantly to measurement errors and limiting the model's reliability in angular deformity classification.

Future improvements should focus on more accurate hip localization, potentially by integrating DICOM metadata, refining templates, and developing a well-annotated dataset with orthopedist-defined landmarks. This would support the evaluation of observer variability and algorithmic accuracy. The high prevalence of hip prostheses in the dataset should be more systematically addressed by developing a dedicated anatomical atlas and a tailored image processing pipeline to accurately handle such cases. While the current approach establishes a functional baseline, further development is needed before clinical application is feasible. Transitioning to deep learning models could offer superior performance by learning robust features directly from diverse radiographic data.

REFERENCES

- Abdallah, Y. M. Y., Alqahtani, T., Abdallah, Y. M. Y., & Alqahtani, T. (2019). Research in Medical Imaging Using Image Processing Techniques. In *Medical Imaging—Principles and Applications*. IntechOpen. <https://doi.org/10.5772/intechopen.84360>
- Arden, N., & Nevitt, M. C. (2006). Osteoarthritis: Epidemiology. *Best Practice & Research Clinical Rheumatology*, *20*(1), 3–25. <https://doi.org/10.1016/j.berh.2005.09.007>
- Bellemans, J., Colyn, W., Vandenneucker, H., & Victor, J. (2012). The Chitranjan Ranawat Award: Is Neutral Mechanical Alignment Normal for All Patients?: The Concept of Constitutional Varus. *Clinical Orthopaedics and Related Research*, *470*(1), 45–53. <https://doi.org/10.1007/s11999-011-1936-5>
- Bellezza, T. (2020, October). Guide to valgus knee prevention. *SelectFlex*. <https://selectflex.com/blogs/how-to-tips/guide-to-valgus-knee-prevention>
- Bercovich, E., & Javitt, M. C. (2018). Medical Imaging: From Roentgen to the Digital Revolution, and Beyond. *Rambam Maimonides Medical Journal*, *9*(4), e0034. <https://doi.org/10.5041/RMMJ.10355>
- Berman, J. J. (2016). Chapter 4—Understanding Your Data. In J. J. Berman (Ed.), *Data Simplification* (pp. 135–187). Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-803781-2.00004-7>
- Bhanage, H., & J., P. (2016). Medical Image Analysis of Image Segmentation and Registration Techniques. *International Journal of Engineering and Technology*, *8*, 2234–2241. <https://doi.org/10.21817/ijet/2016/v8i5/160805205>
- Branch, N. S. C. and O. (2017, March 8). *Osteoarthritis Initiative*. National Institute of Arthritis and Musculoskeletal and Skin Diseases; NIAMS. <https://www.niams.nih.gov/grants-funding/funded-research/osteoarthritis-initiative>
- Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-8*(6), 679–698. <https://doi.org/10.1109/TPAMI.1986.4767851>

- Chen, H., Rogalski, M. M., & Anker, J. N. (2012). Advances in functional X-ray imaging techniques and contrast agents. *Physical Chemistry Chemical Physics*, 14(39), 13469. <https://doi.org/10.1039/c2cp41858d>
- Choudhury, P., Billings, S. J., Bestic, J. M., Peterson, J. J., Stanborough, R. O., Garner, H. W., & Sebro, R. (2023). Radiologists should use the hip-knee-ankle angle rather than the mechanical axis deviation to describe knee alignment. *Skeletal Radiology*, 52(6), 1159–1167. <https://doi.org/10.1007/s00256-022-04234-y>
- Chugh, A. (2024, January 18). MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared—Which Metric is Better? *Analytics Vidhya*. <https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e>
- Cooke, D., Scudamore, A., Li, J., Wyss, U., Bryant, T., & Costigan, P. (1997). Axial lower-limb alignment: Comparison of knee geometry in normal volunteers and osteoarthritis patients. *Osteoarthritis and Cartilage*, 5(1), 39–47. [https://doi.org/10.1016/S1063-4584\(97\)80030-1](https://doi.org/10.1016/S1063-4584(97)80030-1)
- Coxa. (2024). Tekonivelsairaala Coxa Oy. *Tekonivelsairaala Coxa*. Retrieved April 28, 2025, from <https://coxa.fi/coxa-yrityksena/tekonivelsairaala-coxa/>
- CoxaPro—Tiedettä ja tutkimusta Coxassa. (2024). Coxapro. Retrieved June 9, 2025, from <https://coxa.fi/coxapro/>
- Deheyab, A. O. A., Alwan, M. H., Rezzaqe, I. K. A., Mahmood, O. A., Hammadi, Y. I., Kareem, A. N., & Ibrahim, M. (2022). AN OVERVIEW OF CHALLENGES IN MEDICAL IMAGE PROCESSING. *Proceedings of the 6th International Conference on Future Networks & Distributed Systems*, 511–516. <https://doi.org/10.1145/3584202.3584278>
- Deserno, T. (2009). Medical Image Processing. *“Medical Image Processing” in Optipedia, SPIE Press, Bellingham, WA*. <https://spie.org/publications/spie-publication-resources/optipedia-free-optics-information/deserno-medical-image-processing>
- Dhar, T., Dey, N., Borra, S., & Sherratt, R. S. (2023). Challenges of Deep Learning in Medical Image Analysis—Improving Explainability and Trust. *IEEE Transactions on Technology and Society*, 4(1), 68–75. <https://doi.org/10.1109/TTS.2023.3234203>

- Elyan, E., Vuttipittayamongkol, P., Johnston, P., Martin, K., McPherson, K., Moreno-García, C. F., Jayne, C., & Sarker, M. M. K. (2022). Computer vision and machine learning for medical image analysis: Recent advances, challenges, and way forward. *Artificial Intelligence Surgery*, 2(1), 24–45. <https://doi.org/10.20517/ais.2021.15>
- Felson, D. T., Niu, J., Guermazi, A., Sack, B., & Aliabadi, P. (2011). Defining radiographic incidence and progression of knee osteoarthritis: Suggested modifications of the Kellgren and Lawrence scale. *Annals of the Rheumatic Diseases*, 70(11), 1884. <https://doi.org/10.1136/ard.2011.155119>
- Findata. (2025). Findata. Retrieved May 9, 2025, from <https://findata.fi/>
- Foran, J. R. H., & Manner, P. W. (2021, September). *Knee Replacement Implants—OrthoInfo—AAOS*. <https://www.orthoinfo.org/en/treatment/knee-replacement-implants/>
- Galić, I., Habijan, M., Leventić, H., & Romić, K. (2023). Machine Learning Empowering Personalized Medicine: A Comprehensive Review of Medical Image Analysis Methods. *Electronics*, 12(21), Article 21. <https://doi.org/10.3390/electronics12214411>
- Ghuwalewala, S., Kulkarni, V., Pant, R., & Kharat, A. (2021). *Levels of Autonomous Radiology* (No. arXiv:2112.07286). arXiv. <https://doi.org/10.48550/arXiv.2112.07286>
- Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (Fourth edition, Global edition). Pearson.
- Hardy, M., & Harvey, H. (2020). Artificial intelligence in diagnostic imaging: Impact on the radiography profession. *The British Journal of Radiology*, 93(1108), 20190840. <https://doi.org/10.1259/bjr.20190840>
- Heidari, B. (2011). Knee osteoarthritis prevalence, risk factors, pathogenesis and features: Part I. *Caspian Journal of Internal Medicine*, 2(2), 205–212.
- Hirvasniemi, J., Thevenot, J., Immonen, V., Liikavainio, T., Pulkkinen, P., Jämsä, T., Arokoski, J., & Saarakkala, S. (2014). Quantification of differences in bone texture from plain radiographs in knees with and without osteoarthritis. *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society*, 22(10), 1724–1731. <https://doi.org/10.1016/j.joca.2014.06.021>
- Hsu, H., & Siwiec, R. M. (2023). Knee Arthroplasty. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK507914/>

- Huang, J., Tian, F., Zhang, Z., Shi, W., Lin, J., Chen, L., & Yang, H. (2020). Reliability and concurrent validity of angle measurements in lower limb: EOS 3D goniometer versus 2D manual goniometer. *Journal of Orthopaedic Translation*, 24, 96–102. <https://doi.org/10.1016/j.jot.2020.05.002>
- Hunter, D. J., & Bierma-Zeinstra, S. (2019). Osteoarthritis. *The Lancet*, 393(10182), 1745–1759. [https://doi.org/10.1016/S0140-6736\(19\)30417-9](https://doi.org/10.1016/S0140-6736(19)30417-9)
- Insall, J. N., Binazzi, R., Soudry, M., & Mestriner, L. A. (1985). Total Knee Arthroplasty. *Clinical Orthopaedics and Related Research*®, 192, 13.
- Jang, S., Lee, K., & Ju, J. H. (2021). Recent Updates of Diagnosis, Pathophysiology, and Treatment on Osteoarthritis of the Knee. *International Journal of Molecular Sciences*, 22(5), 2619. <https://doi.org/10.3390/ijms22052619>
- Javaid, M., Haleem, A., Singh, R. P., & Ahmed, M. (2024). Computer vision to enhance healthcare domain: An overview of features, implementation, and opportunities. *Intelligent Pharmacy*, 2(6), 792–803. <https://doi.org/10.1016/j.ipha.2024.05.007>
- Jia, Y., Zhai, L., Qin, S., Xu, J., Gao, W., Zhang, B., Wang, X., Zhou, K., Sun, Z., Niu, Y., Bao, H., & Sun, R. (2023). Residual varus alignment after posterior-stabilized total knee arthroplasty limits medial soft tissue remodeling. *BMC Musculoskeletal Disorders*, 24(1), 918. <https://doi.org/10.1186/s12891-023-07048-8>
- Jo, C., Hwang, D., Ko, S., Yang, M. H., Lee, M. C., Han, H.-S., & Ro, D. H. (2023). Deep learning-based landmark recognition and angle measurement of full-leg plain radiographs can be adopted to assess lower extremity alignment. *Knee Surgery, Sports Traumatology, Arthroscopy*, 31(4), 1795. <https://doi.org/10.1007/s00167-022-07124-x>
- Kellgren, J. H., & Lawrence, J. S. (1957). Radiological assessment of rheumatoid arthritis. *Annals of the Rheumatic Diseases*, 16(4), 485–493. <https://doi.org/10.1136/ard.16.4.485>
- Kim, Y.-T., Han, B.-S., Kim, J. B., Sa, J. K., Hong, J. H., Son, Y., Han, J.-H., Do, S., Chae, J. S., & Bae, J.-K. (2024). HKA-Net: Clinically-adapted deep learning for automated measurement of hip-knee-ankle angle on lower limb radiography for knee osteoarthritis assessment. *Journal of Orthopaedic Surgery and Research*, 19(1), 777. <https://doi.org/10.1186/s13018-024-05265-y>

- Kohn, M. D., Sassoon, A. A., & Fernando, N. D. (2016). Classifications in Brief: Kellgren-Lawrence Classification of Osteoarthritis. *Clinical Orthopaedics and Related Research*, 474(8), 1886–1893. <https://doi.org/10.1007/s11999-016-4732-4>
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Laki Sosiaali- Ja Terveystietojen Toissijaisesta Käytöstä, No. 552/2019. <https://www.finlex.fi/fi/lainsaadanto/saaduskokoelma/2019/552>
- Lee, H. S., Hwang, S., Kim, S.-H., Joon, N. B., Kim, H., Hong, Y. S., & Kim, S. (2024). Automated analysis of knee joint alignment using detailed angular values in long leg radiographs based on deep learning. *Scientific Reports*, 14(1), 7226. <https://doi.org/10.1038/s41598-024-57887-1>
- Leong, A. (2023, October 6). Do you know python grayscale histogram analysis? *Medium*. <https://medium.com/@meiyee715/do-you-know-python-grayscale-histogram-analysis-b1f8825aef7d>
- Mahony, N. O., Campbell, S., Carvalho, A., Harapanahalli, S., Velasco-Hernandez, G., Krpalkova, L., Riordan, D., & Walsh, J. (2020). *Deep Learning vs. Traditional Computer Vision* (Vol. 943). <https://doi.org/10.1007/978-3-030-17795-9>
- Manjunath, K. S., Gopalakrishna, K. G., & Vineeth, G. (2015). Evaluation of alignment in total knee arthroplasty: A prospective study. *European Journal of Orthopaedic Surgery & Traumatology*, 25(5), 895–903. <https://doi.org/10.1007/s00590-015-1638-x>
- Manson, E. N., Ampoh, V. A., Fiagbedzi, E., Amuasi, J. H., Flether, J. J., & Schandorf, C. (2019). Image Noise in Radiography and Tomography: Causes, Effects and Reduction Techniques. *Current Trends in Clinical & Medical Imaging*, 3(4), 86–91. <https://doi.org/10.19080/CTCMI.2019.02.555620>
- Mänttari, M. (2023). *Polven tekoniivelpotilaiden radiologinen löydös ennustamassa potilaan raportointia kipua ja toimintakyvyn alenemaa*. <https://trepo.tuni.fi/handle/10024/147799>
- Marques Luís, N., & Varatojo, R. (2021). Radiological assessment of lower limb alignment. *EFORT Open Reviews*, 6(6), 487–494. <https://doi.org/10.1302/2058-5241.6.210015>

- Mayerhöfer, T. G., Pahlow, S., & Popp, J. (2020). The Bouguer-Beer-Lambert Law: Shining Light on the Obscure. *ChemPhysChem*, 21(18), 2029–2046. <https://doi.org/10.1002/cphc.202000464>
- Mikla, V., V.I.Rusin, & Boldizhar, P. (2012). "Advances in Imaging from the First X-ray Images". *Journal of Optoelectronics & Advanced Materials*, 14, 157.
- Nguyen, T. P., Chae, D.-S., Park, S.-J., Kang, K.-Y., Lee, W.-S., & Yoon, J. (2020). Intelligent analysis of coronal alignment in lower limbs based on radiographic image with convolutional neural network. *Computers in Biology and Medicine*, 120, 103732. <https://doi.org/10.1016/j.compbiomed.2020.103732>
- Obuchowicz, R., Strzelecki, M., & Piórkowski, A. (2024). Clinical Applications of Artificial Intelligence in Medical Imaging and Image Processing—A Review. *Cancers*, 16(10), 1870. <https://doi.org/10.3390/cancers16101870>
- OpenCV: Object Detection. (n.d.). Retrieved May 21, 2025, from https://docs.opencv.org/4.x/df/dfb/group__imgproc__object.html#gga3a7850640f1fe1f58fe91a2d7583695dac6677e2af5e0fae82cc5339bfaef5038
- Pei, Y., Yang, W., Wei, S., Cai, R., Li, J., Guo, S., Li, Q., Wang, J., & Li, X. (2020). Automated measurement of hip–knee–ankle angle on the unilateral lower limb X-rays using deep learning. *Physical and Engineering Sciences in Medicine*, 44(1), 53. <https://doi.org/10.1007/s13246-020-00951-7>
- Przystalski, K., Paleczek, A., Szustakowski, K., Wawryka, P., Jungiewicz, M., Zalewski, M., Kwiatkowski, J., Gądek, A., & Miśkowiec, K. (2023). Automated correction angle calculation in high tibial osteotomy planning. *Scientific Reports*, 13(1), 12876. <https://doi.org/10.1038/s41598-023-39967-w>
- Ramazanian, T., Yan, S., Rouzrokh, P., Wyles, C. C., O Byrne, T. J., Taunton, M. J., & Maradit Kremers, H. (2022). Distribution and Correlates of Hip-Knee-Ankle Angle in Early Osteoarthritis and Preoperative Total Knee Arthroplasty Patients. *The Journal of Arthroplasty*, 37(6S), S170–S175. <https://doi.org/10.1016/j.arth.2021.12.009>
- Röntgen, W. C. (1896). On a New Kind of Rays. *Science*, 3(59), 227–231. <https://doi.org/10.1126/science.3.59.227>

- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, 126(5), 1763. <https://doi.org/10.1213/ANE.0000000000002864>
- Schock, J., Truhn, D., Abrar, D. B., Merhof, D., Conrad, S., Post, M., Mittelstrass, F., Kuhl, C., & Nebelung, S. (2020). Automated Analysis of Alignment in Long-Leg Radiographs by Using a Fully Automated Support System Based on Artificial Intelligence. *Radiology. Artificial Intelligence*, 3(2), e200198. <https://doi.org/10.1148/ryai.2020200198>
- Sharda, A. (2021, January 25). Image Filters: Gaussian Blur. *Medium*. <https://aryamansharda.medium.com/image-filters-gaussian-blur-eb36db6781b1>
- Sharma, L., Song, J., Felson, D. T., Cahue, S., Shamiyeh, E., & Dunlop, D. D. (2001). The Role of Knee Alignment in Disease Progression and Functional Decline in Knee Osteoarthritis. *JAMA*, 286(2), 188–195. <https://doi.org/10.1001/jama.286.2.188>
- Sheehy, L., & Cooke, T. D. V. (2015). Radiographic assessment of leg alignment and grading of knee osteoarthritis: A critical review. *World Journal of Rheumatology*, 5(2), 69–81. <https://doi.org/10.5499/wjr.v5.i2.69>
- Sheehy, L., Felson, D., Zhang, Y., Niu, J., Lam, Y.-M., Segal, N., Lynch, J., & Cooke, T. D. V. (2011). Does Measurement of the Anatomic Axis Consistently Predict Hip-Knee-Ankle Angle (HKA) for Knee Alignment Studies in Osteoarthritis? Analysis of long limb radiographs from the Multicenter Osteoarthritis (MOST) Study. *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society*, 19(1), 58–64. <https://doi.org/10.1016/j.joca.2010.09.011>
- Simon, S., Schwarz, G. M., Aichmair, A., Frank, B. J. H., Hummer, A., DiFranco, M. D., Dominkus, M., & Hofstaetter, J. G. (2022). Fully automated deep learning for knee alignment assessment in lower extremity radiographs: A cross-sectional diagnostic study. *Skeletal Radiology*, 51(6), 1249–1259. <https://doi.org/10.1007/s00256-021-03948-9>
- Tack, A., Preim, B., & Zachow, S. (2021). Fully automated Assessment of Knee Alignment from Full-Leg X-Rays employing a “YOLOv4 And Resnet Landmark regression Algorithm” (YARLA): Data from the Osteoarthritis Initiative. *Computer Methods and Programs in Biomedicine*, 205, 106080. <https://doi.org/10.1016/j.cmpb.2021.106080>

- Tanner, I. L., Ye, K., Moore, M. S., Rechenmacher, A. J., Ramirez, M. M., George, S. Z., Bolognesi, M. P., & Horn, M. E. (2024). Developing a Computer Vision Model to Automate Quantitative Measurement of Hip-Knee-Ankle Angle in Total Hip and Knee Arthroplasty Patients. *The Journal of Arthroplasty*, 39(9), 2225–2233. <https://doi.org/10.1016/j.arth.2024.04.062>
- Tapio, R. (2025). Vuosikertomus 2024. *Tekonivelsairaala Coxa*. <https://coxa.fi/vuosikertomus/THL>. (2024, September 25). *Tekonivelleikkaukset 2023*. https://repo.thl.fi/sites/tekonivelleikkaukset/tekonivelleikkaukset_2023_liitetaulukot.html#taulukko3a
- Wang, J., Hall, T. A. G., Musbahi, O., Jones, G. G., & Van Arkel, R. J. (2023). Predicting hip-knee-ankle and femorotibial angles from knee radiographs with deep learning. *The Knee*, 42, 281–288. <https://doi.org/10.1016/j.knee.2023.03.010>
- Wirtanen, M. (2024). *Alaraajan mekaaninen akselimittaus*. https://huslab.fi/radiologia/02_tutkimukseen_lahettaminen_ajanvaraus_ja_esivalmistelu/natiivitutkimukset/04_radiologisten_tutkimusten_menettelyohjeet/07_alaraajojen_natiivitutkimus/alaraajan_mekaaninen_akselimittaus.pdf
- Yan, S., Ramazanian, T., Chaudhary, V., & Maradit Kremers, H. (2022). Deep Learning Method for Hip Knee Ankle Angle Prediction on Postoperative Full-Limb Radiographs of Total Knee Arthroplasty Patients. *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 5070–5073. <https://doi.org/10.1109/EMBC48229.2022.9870936>
- Zorzi, A. R., & Miranda, J. B. D. (Eds.). (2018). *Primary Total Knee Arthroplasty*. InTech. <https://doi.org/10.5772/intechopen.68269>