INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

ROBERT GORDON UNIVERSITY ABERDEEN

Knee Osteoarthritis Detection Using Machine Learning on Pre-Processed X-ray Images and Treatment Recommendation

Group 18 Thesis by

Nuha Rameez- IIT: 20210892, RGU: 2236762

Dewmi Dissasnayaka- IIT: 20220714, RGU: 2313079

Chanula Gajadeera Arachchige- IIT: 20220609, RGU: 2236763

Minudaka Liyanage- IIT: 20210902, RGU: 2236755

Supervised by

Ms. Vishmi Embuldeniya

Submitted in partial fulfillment of the requirements for the BSc (Hons) Artificial Intelligence and Data Science degree at the Robert Gordon University.

March 2024

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Chapter Overview

The multiple machine learning model application proposed, will detect and recommend non-medical treatments for osteoarthritis in knee joints. As there are no proper knee osteoarthritis detection systems currently implemented, most medical professionals must manually grade the severity which is considered a competitive task. The aim of project is to ease the task of classifying the severity grade of the osteoarthritis from X-ray images, for the medical professionals and guide them in giving recommended treatments.

1.2 Problem Domain

Knee osteoarthritis is a degenerative joint disorder that affects millions of people worldwide, which significantly lowers their quality of life. In response to the aging population and the increasing prevalence of obesity, knee osteoarthritis is on the rise around the world. To address this growing health concern, it is necessary to develop a Knee Osteoarthritis Detection System.

A few key factors contribute to the significance of knee osteoarthritis detection system research. First, early detection and intervention are essential for addressing knee osteoarthritis, as it enables pretreatment and disease prevention. Second, the system will categorize the degree of the disease according to the Kellgren Lawrence System, allowing the user to receive treatment for the relevant grade. Thirdly this system recommends treatments for every grade.

The magnitude of the problem addressed by the Knee Osteoarthritis Detection System is significant and complex. This system has a broad scope, utilizing advanced image processing and machine learning techniques to analyze x-ray images of knees. By doing so, it can determine whether a knee is affected by the disease or not. In cases where a knee is found to be diseased, the system further categorizes the level of severity, classifying it as a healthy knee, doubtful, minimal, moderate, and severe. With this information, the automated knee osteoarthritis detection system then provides tailored treatment recommendations corresponding to the relevant level of the disease. On the other hand, if a knee is found to be in a healthy state, the system recommends precautionary measures.

1.3 Problem Definition

A large percentage of the population suffers from the musculoskeletal condition known as knee osteoarthritis. Due to aging-related wear and tear and a progressive loss of articular cartilage, this is a very common condition found in the elderly. Knee osteoarthritis is usually a progressive illness with the potential to make patients disabled over time. The intensity of the disease and the rate of progression differ from person to person. This disease is usually treated using conservative methods, however in worst case scenarios, patients may require surgery. (Hsu and Siwiec, 2023) For prompt action and efficient patient care, knee osteoarthritis must be accurately and early diagnosed. The probability of treating osteoarthritis on the knee without requiring the patient to undergo surgery increases with early diagnosis. Conventional diagnosis techniques, which frequently depend on examining X-ray pictures manually, can be time consuming and biased. Moreover, healthcare practitioners face difficulties due to the growing amount of medical imaging data. By using machine learning approaches, this study seeks to solve the issue of fast and accurate knee osteoarthritis detection and recommending suitable treatment plans.

1.4 Research Motivation

Osteoarthritis is a common progressive joint condition which was calculated to be present in 528 million people in 2019 (World, 2023). The knee being the most recurrently affected joint, it was the perfect target for this research. As early detection of knee osteoarthritis is beneficial in avoiding more momentous orthopedic issues, with this detection system doctors can easily carry on with their osteoarthritis procedure and treat the patient accordingly. With the prime motivation for this research being able to detect knee osteoarthritis faster and easier than a doctor would, making this research a certain success.

1.5 Existing Work

Table 1Existing Work

Citation	**Technology/ Algorithm**	**Advantage**	**Limitation**
(Kotti et al., 2017)	Random Forests	Automatically determines the extent of knee osteoarthritis by identifying patterns that are more indicative of the condition.	Not validated using radiograph images (ie; X-rays)
(S. Sheik Abdullah & M. Pallikonda Rajasekaran, 2022)	Kellgren-Lawrence (KL) grading system and deep learning	After inserting the x-ray, the system will automatically grade it according to Kellgren-	Not recommending precautions for the non-diseased knees.

		Lawrence grading value	
(Tiulpin A, Thevenot J, Rahtu E, Lehenkari P, Saarakkala S. Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach.)	Deep Siamese Convolutional Neural Network	The model's capacity to extract pertinent features from OA that can be applied to diverse datasets.	It is not possible to merge disparate medical datasets.
(Chen, P., Gao, L., Shi, X., Allen, K. and Yang, L. 2019)	CNN model named YOLOv2	Comparison of different CNN models for knee osteoarthritis detection	Has to manually annotate knee joints in x-rays
(KOKKOTIS, C. et al., 2020.)	Linear Regression	Easy to implement	Overly simplistic for capturing intricate relationships between variables, leading to a susceptibility to overfitting.

1.6 Research Gap

The systems that currently exist have osteoarthritis detection and treatment recommendation as separate entities, resulting in a gap that needs to be bridged to integrate them into a unified approach.

Our project aims to bridge this gap by integrating osteoarthritis detection, severity assessment, and personalized treatment recommendations into a unified and holistic system. This integrated approach is designed to elevate the overall quality of patient care, streamline healthcare decision-making processes, and ultimately enhance patient outcomes.

Through the merging of these traditionally separate components, our research endeavors to make a significant contribution to the advancement of healthcare technology and offer a more comprehensive and effective solution for managing the complexities of knee osteoarthritis.

1.7 Contribution to the Body of Knowledge

1.7.1 Technological Contribution

This Knee Osteoarthritis detection system will be running as a web application, created using HTML, CSS, and JavaScript. TensorFlow will be used to build the Convolutional Neural Network model which will check if the X-ray is of a knee bone or not and the Binary classification model which will detect whether osteoarthritis is present in the joint. SciKit-Learn and TensorFlow will be used to build a custom neural network model which classifies the severity of the osteoarthritis. Scikit-learn will be used to build Random forests models which will recommend the suitable treatment plans accordingly. Python programming language and the required libraries will be used to implement this system.

1.7.2 Domain Contribution

Using all the mentioned technologies, this system will assist the medical field of musculoskeletal health with a method to detect knee osteoarthritis comparatively easier to normal methods.

Knee Osteoarthritis, if detected early, will be advantageous for patients as they can get treatment methods to slow down the process of osteoarthritis. With this system, the detection of osteoarthritis is made easier, where a patient will only need an X-ray of their knee joint to diagnose whether osteoarthritis is present or not. Normal diagnosis of orthopedic surgeons/rheumatologists will take more time since early osteoarthritis does not usually appear in X-rays (NIAMS, 2017). With a deep understanding of both traditional and deep learning approaches, we create algorithms capable of detecting subtle signs of osteoarthritis, distinguishing them from normal variations and other pathologies, and predicting disease progression.

Finally, the system will recommend treatment plans which are based on the feedback of orthopedic surgeons. It's taken according to the severity of the osteoarthritis, and how the surgeons would normally recommend treatments according to the specific severity. This system ultimately makes the process of detecting Knee Osteoarthritis easier for medical professionals.

1.8 Research Challenge

- One major challenge is to find a sufficiently large and high-quality collection of pre-processed knee X-ray images with precise annotations for osteoarthritis in the knee. Acquiring those data sets to train a model was time consuming.
- Finding research papers is a major challenge as currently there is no system that utilize all four components such as CNN based knee feature extraction, Binary classification for osteoarthritis detection, a model for severity classification and random forest-based treatment recommendations in use; especially research on treatment recommendation

system integrated with the diagnosing system was notably scarce, showing a gap in existing literature.

- Removing noise and changes in image quality is a crucial step when cleaning, standardizing, and processing X-ray images. It can be challenging to find consistent data pre-processing techniques for different types of X-ray images considering the time in hand for this project.
- Extracting relevant features from a knee x-ray to precisely capture the osteoarthritis and finding the most suitable algorithms to identify them in the given short period of time and while trying out different algorithms that has potential, might be a challenge as it directly connects to the model performance and time-consuming nature of the system.
- Developing a robust and personalized non-medical treatment recommendation system that takes condition's severity and characteristics into consideration (and giving necessary precautions to avoid osteoarthritis) is a complex task. Also validating systems performance in real clinical settings and ensuring patients outcomes has been improved, necessitates close collaboration with healthcare professionals. This poses a variety of challenges for this research project.

1.9 Research Questions

RO1: Which models are best for reading x-rays, checking if the joint is normal or not, determining the severity of a diseased joint and recommending suitable treatment plans?
R02: How will the model detect the x-ray uploaded is an x-ray of the knee joint?
R03: How will the system determine the severity of osteoarthritis from x-ray?

1.10 Research aim

This proposed system uses a machine learning system to accurately detect knee osteoarthritis using X-ray images that are pre-processed and subsequently provide a personalized treatment recommendation to the patient according to the severity of the condition. This system will help the early diagnosis and management of knee osteoarthritis, improving patient outcomes, and reducing strain on the doctors.

1.11 Research Objective

Table 2Research Objectives

Research Objectives	Explanation	Learning Outcome
Problem Identification	RO1: Identifying a way to accept only knee joints from X rays entered to the system.	LO1

	RO2: Identify the most suitable CNN model to Extract the knee joint from the x-ray. RO3: Check if the knee joint in the x-ray is normal or diseased. RO4: Implementing a system that can identify severity of osteoarthritis. RO5: Integrating a treatment recommendation system into the system.	
Literature Review	The following areas were the focus of literature search. RO1: Identifying the methods used to identify and extract features of the knee bone. RO2: Identify the existing methods that are used to determine whether knee bone has osteoarthritis. RO3: Study how severity of the disease is detected. RO4: Research on providing treatment plans according to the severity of osteoarthritis.	LO1
Data Gathering and Analysis	• Knee Osteoarthritis Dataset with “Severity Grading" from Kaggle. • Interviewing specialists in the field of knee osteoarthritis • Data gathered from Research papers available through IEEE, Google scholar, ScienceDirect etc.	LO2, LO3
Research Design	Observational Research Design: Analyze existing datasets of X-ray images and patient data to investigate patterns and associations related to knee	LO3, LO4

	osteoarthritis. Useful for understanding relationships without direct experimental manipulation.	
Implementation	Implementation of different parts of this project as follows: ● Deploy a CNN model that identifies and extracts data from an x-ray of a knee bone. ● Binary classification to check presence of osteoarthritis in knee joint. ● Use of custom neural network model to check severity of osteoarthritis. ● Develop a treatment recommendation system according to severity index using Random Forests.	LO2, LO3, LO4
Testing and Evaluation	X-ray image datasets have been gathered separately for both testing and evaluating the system. To increase the system's accuracy on making decisions, it will be tested and examined by using those data sets.	LO2, LO4

1.12 Project Scope

1.12.1 In-scope

Table 3Project In-Scope

No	Description
01	Checking whether the x-ray input is a knee-based x-ray or not.
02	Predicting whether the user input x-ray is normal or has osteoarthritis.
03	Classification of the knee with osteoarthritis according to the severity

04	Recommending non-medical treatments for the diseased knee and precautions for the non-diseased knee.

1.12.2 Out-scope

Table 4Project Out-Scope

No	Description
01	Detecting the x-ray image and categorizing to which part of the body the x-ray image belongs to.
02	Predicting and categorizing the user input x-ray according to the type of disease.
03	Classify each disease according to the severity.
04	Recommending both medical and physical treatments for the diseased bones and precautions for the non-diseased bones.

1.12.3 Prototype Diagram

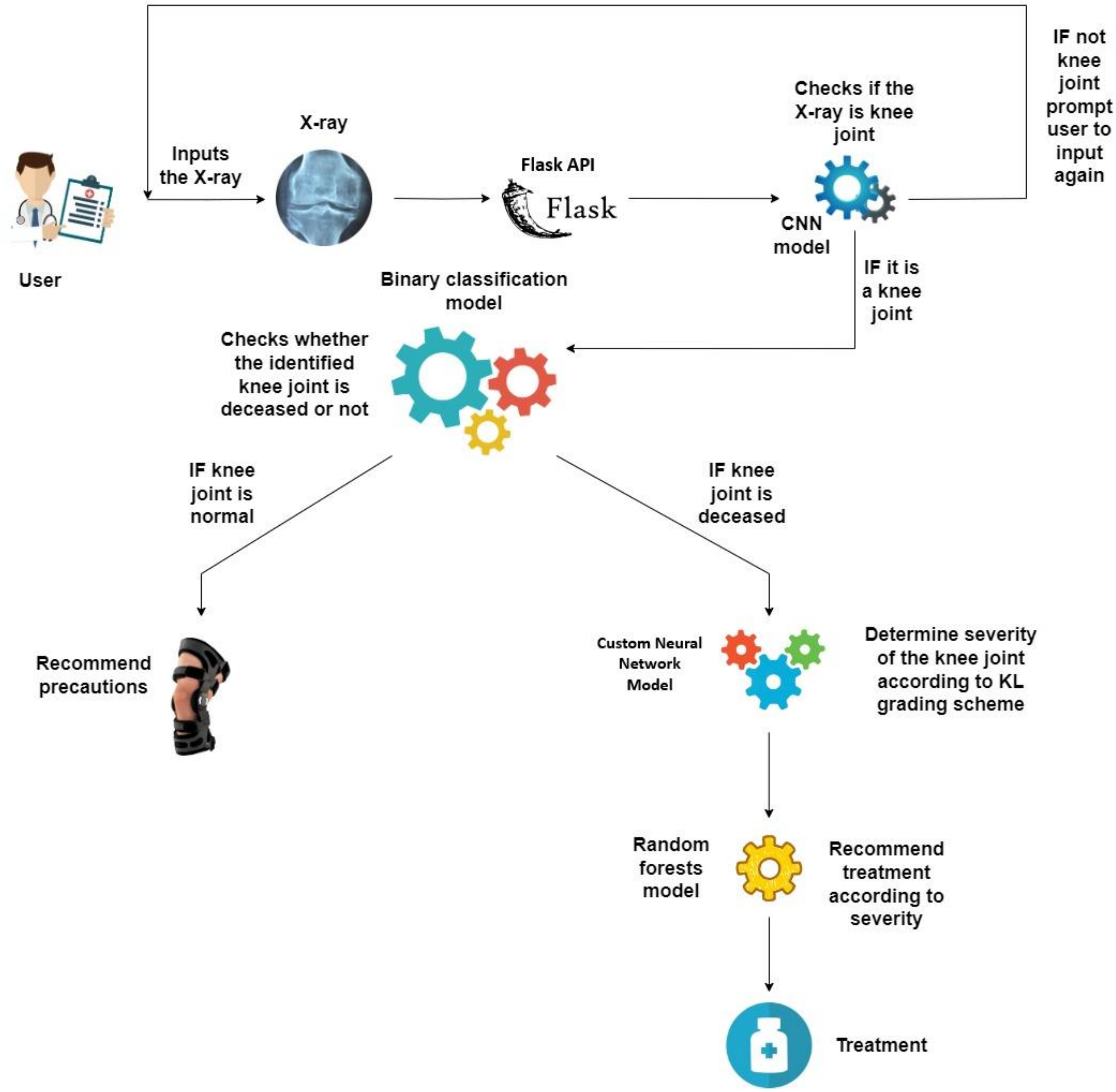

Figure 1Prototype Design

1.13 Resource requirements

1.13.1 Hardware requirements

A modern laptop or desktop with a multi-core processor (Intel Core i5 or i7) and with a clock speed of 2.5 GHz or higher as these are suitable for many machine learning applications and can handle training and inference processes more efficiently. RAM higher than or equal to 16GB is required to work with moderately sized datasets and train models. 256GB or higher SSD would be sufficient as it can hold the dataset and run the system.

1.13.2 Software requirements

- Python – Python is the primary language proposed to develop the system since it has library support and is excellent in error handling.

- PyCharm- Utilized in the creation of commercial and proprietary software.
- TensorFlow – To train and to pre-process the CNN models.
- Scikit-learn – A python library utilized for machine learning and statistical modeling including classification, regression and clustering
- Flask- Flask is used to create web applications and APIs that provide an easy to use and adaptable framework for managing HTTP requests and responses.
- HTML- Utilized for creating the webpage interface.
- CSS – Utilized for web page layout and style.
- JavaScript – Utilized for scripting the webpage.
- React – Used to create dynamic and responsive web interfaces.
- MS Word – Utilized for documentation purposes of the research.
- Windows Operating System – To manage computational functionalities and host the development environment.

1.13.3 Data requirements

Since a machine learning model will be trained to detect knee osteoarthritis, the plan is to utilize Knee X-ray datasets with severity grading that are readily available on Kaggle. In addition to that past research that is connected to this domain will be referred.

1.13.4 Skills requirements

- Programming skills
- Computer Vision Knowledge
- Deep learning and Machine learning Skills
- Web development
- API development
- Problem Solving
- Time management

1.14 Chapter Summary

Osteoarthritis being a disease that cannot be directly diagnosed by a human in an instant, leads to diagnosis of it in the late stages. Therefore, it is important to have an application that can instantly classify osteoarthritis effectively. Identifying the severity of the knee osteoarthritis accurately and recommending suitable treatment recommendation allows to bring down the rate of permanent disability due to knee osteoarthritis in patients drastically.

Chapter 2: Literature Review

2.1 Chapter Overview

Osteoarthritis is one of the most common forms of arthritis, which affects several people worldwide. This mostly occurs on the joints of hands, knees, hips and spine. This project focuses on osteoarthritis that occurs in knees. Studies show that more than 80% of adults over 55 years of age suffer with this disease (Clinic, 2023). There are several ways to detect osteoarthritis in knees, this project focuses on detecting it from x-rays. The purpose of this project is to computerize the detection process. To implement this, this project uses four AI components to simplify and increase efficiency of the system. The components used in this project are CNN- Based Knee Region Extraction, Binary Classification for Osteoarthritis Detection, Custom Neural Network for Severity Classification and Random Forest based System for Treatment Recommendation. This system will not only detect osteoarthritis in knee, but also recommend suitable precautions and treatments according to the severity of the disease.

2.2 Concept map

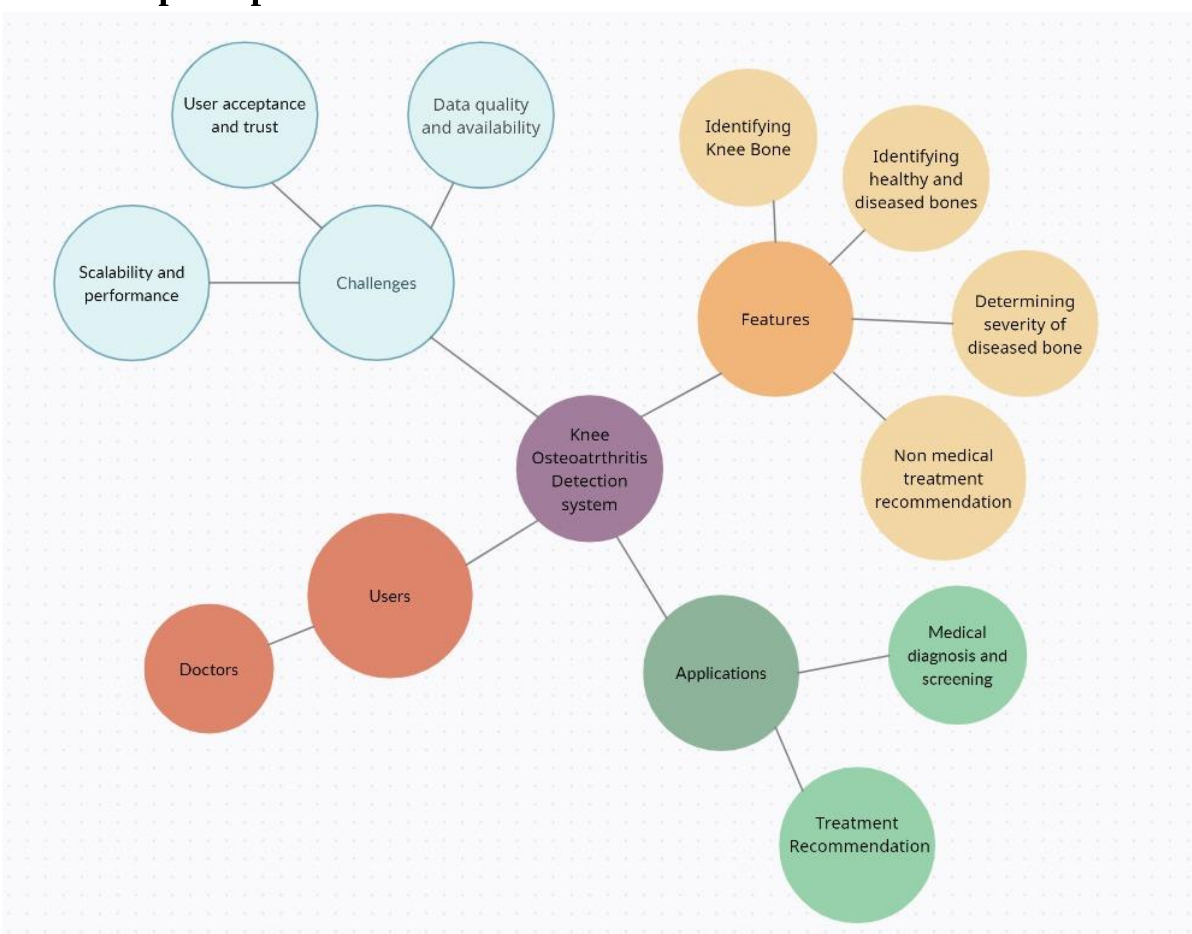

Figure 2 Concept Map

2.3 Existing Work

Table 5 Existing Work (LR)

Research	Author	Year	Dataset	Model Used	Metric
CNN-Based Knee Region Extraction					
Automatic Detection of Knee Joints and Quantification of Knee Osteoarthritis Severity Using Convolutional Neural Networks	Joseph Antony, Kevin McGuinness, Kieran Moran, Noel E. O'Connor	2017	bilateral PA fixed flexion knee X-ray images from the Osteoarthritis Initiative (OAI) and Multicenter Osteoarthritis Study (MOST) in the University of California, San Francisco	CNN, FCN	CNN accuracy 63.4%
Knee Osteoarthritis Detection Using Deep Feature Based on Convolutional Neural Network	Dilovan Asaad Zebari, Shereen Saleem Sadiq, Dawlat Mustafa Sulaiman	2022	from well-known local hospitals and health centers to evaluate, including 2000 computer-generated knee X-rays	CNN,SVM, KNN	90.01% Accuracy
Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks	Joseph Antony, Kevin McGuinness, Kieran Moran, Noel E. O'Connor	2016	bilateral PA fixed flexion knee X-ray images, taken from the baseline (image release version O.E.1) radiographs of the	a linear SVM and the Sobel horizontal image gradients as the features for detecting the knee joint centers.	a linear SVM produced 95.2% 5-fold cross validation and 94.2% test accuracies.

			Osteoarthritis Initiative (OAI) dataset containing an entire cohort of 4, 476 participants	CNN for assessing the severity of knee OA through classification and regression.	
Binary Classification for Osteoarthritis Detection					
Machine Learning Approaches for the Classification of Knee Osteoarthritis	Tayyaba Tariq, Zobia Suhail, Zubair Nawaz	2023	OsteoArthritis Initiative (OAI)	SVM, LR, SGD, KNN, MLP, Perceptron	LR accuracy (most accurate model): 84.5%
Estimating the severity of knee osteoarthritis using Deep Convolutional Neural Network based on Contrast Limited Adaptive Histogram Equalization technique	Amina A. Abdo, Wafa El-Tarhouni, Asma Fathi Adbulsalam, Abdelgader Bubaker Altajori	2022	Knee Osteoarthritis severity grading dataset (Dataset I) Osteoarthritis Initiative (OAI) (Dataset II)	Binary Classification	Dataset I: 83.5% Dataset II: 85.5%
A fully automatic fine tuned deep learning model for knee osteoarthritis detection and progression analysis	Sameh Abd El-Ghany, Mohammed Elmogy, A. Abd El-Aziz	2023	OsteoArthritis Initiative (OAI)	DenseNet169, InceptionV3, Xception, ResNet50, DenseNet121, InceptionResNetV2	DenseNet169 (most accurate model): Accuracy: 93.78% F1-score: 89.27%
Identifying Severity Grading of Knee Osteoarthritis from X-ray Images Using an Efficient	Sozan Mohammed Ahmed, Ramadhan J. Mstafa	2022	OsteoArthritis Initiative (OAI)	DHL-II	Accuracy rate: 90.8%, Average F1-score: 90.5%

Mixture of Deep Learning and Machine Learning Models					
Custom Neural Network for Severity Classification					
Classification of knee osteoarthritis based on quantum-to-classical transfer learning.	Dong Yumin, Che Xuanxuan, Fu Yanying, Liu Hengrui, Zhang Yang, Tu Yong	2023	Dataset from Chen et al Quinonero-Candela et al.	A hybrid quantum convolutional network	Accuracy (98.36%), Validation Accuracy (56.38%)
Automatic Detection and Classification of Human Knee Osteoarthritis Using Convolutional Neural Networks	Mohamed Yacin Sikkandar1,*, S. Sabarunisha Begum2, Abdulaziz A. Alkathiry3, Mashhor Shlwan N. Alotaibi1 and Md Dilsad Manzar4	2021	X-ray images of 350 subjects at Durma and Tumair General Hospital, Riyadh	Deep Siamese Convolutional Neural Network	Accuracy (93.2%) Validation Accuracy (93.2%)

Transfer Learning-Based Smart Features Engineering for Osteoarthritis Diagnosis From Knee X-Ray Images	Amjad Rehman, Ali Raza, Faten S.Alamri, Bayan Alghofaily, Tanzila Saba	2023	Dataset based on 3,000 subjects with 5,960 knees image from the osteoarthritis initiative	Convolutional Neural Network	Accuracy (99%)
Deep Neural Network-based Knee Osteoarthritis Grading Using X-Rays	Zainab Mushtaq Wani, Dr. Satish Saini	2022	Rheumatology Initiative 1	YOLOv2 model to identify the knee joint and fine-tune CNN models	Accuracy: 96.7%
Random forest System for Treatment Recommendation					

Risk controlled decision trees and random forests for precision Medicine	Kevin Doubleday, Jin Zhou, Hua Zhou, Haoda Fu	2021	DURABLE dataset	Decision tree and random forest	rcDT to 86.4% and rcRF was 92.1%
Ensemble Machine Learning (Grid Search & Random Forest) based Enhanced Medical Expert Recommendation System for Diabetes Mellitus Prediction	Nagaraj P, Muneeswaran V, Muthamil Sudar K, Naga Vardhan Reddy A, Deshik G, Charan Kumar Reddy C	2022	Mendeley datasets	Grid search and random forest algorithms	accuracy of 98%
A Recommendation System for Diabetes Detection and Treatment	Fatima Almatrooshi, Sumayya Alhammadi, Said A. Salloum, Iman Akour, Khaled Shaalan	2020	TunedIT.org	random forest	79.2% accuracy
An algorithm for generating individualized treatment decision trees and random forests	Kevin Doubleday, Hua Zhou, Haoda Fu and Jin Zhou	2018	IPWE and AIPWE	Decision trees and random forest	ITR for Scheme A.1 99.7%, A.2 86.3%, B.1 99.7%, B.2 86.5%, C.1 94.3%

2.4 Technology/Approach/Algorithm Review

1. CNN- Based Knee Region Extraction

Osteoarthritis occurs in the joints of hands, knees, hips, and spine. Our system focuses on detecting Osteoarthritis on knees only. Therefore, this model will be used to check if the x-ray input into the system is of a knee. This model also focuses on feature extraction and improving the quality of the x-ray to allow the other models to work smoothly.

After the x-ray is uploaded it needs to be preprocessed before the data can be retrieved.to accomplish this in research done by Zebari, Sadiq and Sulaiman (2022), . Researchers have used a filter named Finite Impulse Response (FIR) which enhances the quality of knee trabecular texture.

Despite attempts to ensure consistency, X-ray images may vary from patient to patient, which causes the joint position of each x-ray to be different. To overcome this, research done by Joseph, Kevin, Kieran, Noel (2017) have used a fully convolutional neural network to detect ROI (region of interest) knee joint in a knee x-ray. The model incorporates 4 stages of convolutional layers followed by a max pooling layer, which allows to increase the clarity and quality of the x-rays.

Next CNN is used for feature extraction in order to get the needed information from the knee x-ray. In the research done by Zebari, Sadiq and Sulaiman (2022), 2 dimensional CNN was used like most of the research in this field. This architecture has multiple convolutional layers which take the image as input and the filters are performed on it. Among the several CNN feature classifiers used, KNN classifiers had the highest accuracy of 90.10%.

Also it is important to note that another research done by Joseph, Kevin, Kieran and Noel(2016) used linear SVM and Sobel horizontal image gradings in order to detect knee joints which helps with diagnosing osteoarthritis. Pre-trained CNNs were also used for feature extractions in knee x-rays in their study.

Another research done to compare the accuracy between the machine and the surgeon by Schwartz et al., (2020) states, CNN models were used to read x-rays. This shows us CNN data extraction decisions were more related to a surgeon’s decision.

This model allows us to confirm if the x-ray entered into the system is an x-ray of a knee joint, as our system focuses only on knee osteoarthritis detection.

2. Binary Classification for Osteoarthritis Detection

After the X-ray uploaded into the system is confirmed to be an X-ray of the knee joint, this model is used to determine if the knee in the x-ray is normal or is affected by osteoarthritis. Since the outcome will be one of two possible results, a binary classification model has been used for this purpose.

Many researchers have used binary classification for this same purpose as well. According to research conducted by Tariq, Suhail and Nawaz (2023) many researchers have combined the different grades of osteoarthritis (KL classes) of a dataset to make two classes and use this as a dataset to conduct binary classification. An example used to prove this is Bayramoglu et al. (2020) divided their dataset into two classes and used a Logistic Regression classifier to conduct the classification. Furthermore, it is recorded in this research that six experiments were performed to determine the best model. The experiments were conducted by a Linear Support Vector Machine model, Logistic Regression model, Stochastic Gradient model, Perceptron, K Nearest Neighbors, Artificial Neural Network-based Multi-layer Classifier. After testing and validating, it was proven that the Logistic Regression had a higher accuracy in comparison to other models. The accuracy average was 84.5% (Tariq, Suhail and Nawaz, 2023).

The research conducted by Abdo et al. (2022) showed that there was excellent accuracy when using binary classification. In the two datasets used in this scenario, the binary classification model produced an accuracy of 83.50% and 85.50% respectively. Sameh Abd El-Ghany, Elmogy and A. A. Abd El-Aziz (2023) used DenseNet169 to propose a model and compared it to the five Deep Learning approaches, which were InceptionV3, Xception, ResNet50, DenseNet121 and IndceptionResNetV2. Here, when training, the dataset was divided into two classes, one with osteoarthritis and one without osteoarthritis. Finally, they concluded that in binary classification, DenseNet169 had an accuracy of 93.78% and a F1-score of 89.27%. Another study by Sozan Mohammed Ahmed and Mstafa (2022) mentions that the primary focus of the study in classifying knee osteoarthritis was binary classification. Here, they too determined if the x-ray of the knee joint was normal or if it was affected by osteoarthritis. For this scenario, they proposed a DHL-II model and it generated an accuracy of 90.8% and an F1-score of 90.5% respectively.

It is important to note that all the above-mentioned research used several other models other than binary classification as well. However, as per this feature's necessity, which is to only detect if the knee bone is normal or not, only the research regarding the use and accuracy of the binary classification model has been included in this section. The purpose of this feature is to reduce complexity when it comes to treatment and precaution recommendation. If the bone is normal, the system will directly access the model which will suggest precautionary measures (2.4), however, if the knee is suffering with osteoarthritis, it will access the model which will determine the severity of the disease (2.3) first and then direct to (2.4) to get treatment recommendations.

3. Custom Neural Network for Severity Classification

Neural networks are a machine learning approach which is popularly used in diagnosing knee osteoarthritis. It is used to predict the severity of osteoarthritis of patients by following the Kellgren-Lawrence Classification System.

There are many research which have used neural networks for classification of osteoarthritis. The research conducted by DONG, Y. et al., 2023 uses a hybrid quantum convolution network structure to test for osteoarthritis in knee images dataset. The results of the experiment show that the quantum classical transfer learning approach they took can efficiently classify the severity with an accuracy of 98.36%.

In another research conducted by Mohamed Yacin Sikkandar 2024, it utilizes an unsupervised segmentation algorithm, the local center of mass, streamlines the isolation of the tibiofemoral joint (ROI) which all helps to accurately extract features. The resulting features extracted; the neural network achieved an accuracy of 93.20%.

Research by Amjad Rehman, 2024 uses transfer learning-based feature engineering, which uses pre-trained neural networks to extract relevant features from images, along with hyperparameter

optimization and cross-validation, on Convolutional Neural Networks model to give an experimental accuracy of 99%.
Another research Zainab Mushtaq Wani, IJRASET, 2024, which uses a customized YOLOv2 network to identify knee joints in x-rays with low contrast, which is then used with popular convolution neural networks such as ResNet, VGG, DenseNet and InceptionV3. When evaluated the models with the datasets, an accuracy of 96.7% and a mean absolute error of 0.344 is obtained for KL grading.

4. Random Forests Based System for Treatment Recommendation

After confirming the grade of the disease from the relevant x-ray, this model is used to recommend treatments and precautions. Random forests will be used to implement this recommendation system.

Many researchers have used random forests for different sectors. According to the research conducted by Nagaraj p et al (2022) a random forest generates accurate predictions that are simple.. In comparison to the decision tree method, the random forest algorithm is more accurate at predicting outcomes. To prove that Nagaraj p et al (2022) has done research on the diabetes disease using the Mendeley dataset. The experiments were conducted using grid search, XG boost, and random forest. After testing it was proven that the random forest has the highest accuracy with 98% and 93% for the precision. S. S. Bhat and G. A. Ansari (2021) also has done similar research on diabetes treatment recommendation systems.

The research conducted by Kevin Doubleday,Jin Zhou, Hua Zhou, Haoda Fu (2021) has done a comparison on the random forest and decision trees. The work proposed two methods of risk controlling such as a risk-controlled decision tree (RCDT) and risk controlled random forests (RCRF). In this experiment they have categorized the research into 5 categories. In the discussion of this research, they have concluded that the random forest model is the safest and easiest to understand with an accuracy of 92.1%, and the decision tree with a 86.4% accuracy. This work contributes to deciding proper treatment rules based on both efficiency and risk considerations.

The research conducted by Kevin Doubleday et al (2018) has mainly focused on treatmenting diabetes disease using decision trees and random forest models. Here, they have proven that the random forest model can do the treatment for subcategories as well. They have categorized their dataset into sections as A.1, A.2, B.1, B.2, C.1, C.2. Accuracy for each level is concluded as 99.7%, 86.3%, 99.7%, 86.5%, 94.3% respectively.

In conclusion this treatment recommendation system is used to treat each and every level of osteoarthritis. It also includes that this system provides precautions for the non-diseased knees as well.

2.5 Tools and Techniques

A robust stack of tools and techniques that can be used to successfully implement this project are as follows:

- Python: a high-level programming language commonly used in machine learning.
- PyCharm: an Integrated Development Environment for Python programming. This provides features such as completion, debugging and project management.
- TensorFlow: an open-source machine learning framework. This allows implement several machine learning concepts proposed in this project.
- Scikit-learn: another machine learning library that provides simple and efficient tools for data analysis and modeling.
- React: a JavaScript library for building user interfaces.
- Flask: a lightweight web framework for Python, which allows to build web applications easily by providing the necessary tools.
- Git: a version control system that allows to track changes in code, collaborate with team members and manage different versions of the project.
- GitHub: a web-based platform that hosts Git repositories and provides additional features for collaboration.

2.6 Chapter Summary

The purpose of this project is to apply machine learning concepts on preprocessed knee x-rays to detect osteoarthritis. This project will first focus on identifying the input X-ray image to identify whether it's a knee bone or in order to improve the accuracy of diagnosing osteoarthritis. This project will not only detect knee osteoarthritis but will detect the degree of the disease and provide suitable treatment plans as well. Furthermore, this project also focuses on providing precautionary measures to normal knees. Several machine learning concepts such as CNN, Binary Classification, Custom Neural Network and Random Forests will be used to implement this project.

Chapter 3: Methodology

3.1 Chapter overview.

This chapter outlines the methodologies used for research, development and project management. Several approached were utilized to suit specific requirements of the project. The chapter also offers a thorough analysis of the decision-making procedures that led to selecting the respective approaches.

3.2 Research methodology.

Table 6 Research Methodology

Research Philosophy	A positivism research philosophy for knee osteoarthritis detection from X-rays has been used for this project. It prioritizes objective, data-driven approaches, using extensive datasets, machine learning, and statistical analysis to develop a highly accurate diagnostic system.
Research Approach	This research uses a quantitative approach since the dataset comprises numerical data from X-ray images, which allows to precisely quantify the performance of the machine learning models in detecting knee osteoarthritis.
Research Strategy	Experimental Design research strategy has been used for this project as for knee osteoarthritis detection this method guarantees controlled tests, thorough data analysis and dependable cause-and-effect results.
Research Choice	This research will be a mono method as it relies solely on one data-gathering method. This focused approach may be used when depth of analysis, resource constraints, or specific research objectives warrant a single data-gathering method.
Time zone	Cross-sectional time zone has been used since a fixed number of datasets has been used in this project for the purpose of training, testing, and validating. This also aligns with this project's goal to assess the effectiveness of this system at a single time point.

3.3 Development methodology.

a. The iteration model has been selected as the life cycle model for this project. This project prioritizes the evolving requirements, thus seeing the need to keep adapting to new insights.

An iterative model such as Agile allows this project to respond to changing research objectives. This model also allows space for continuous improvement which enhances accuracy and effectiveness. This also allows to identify and mitigate risks, accommodates experimentation, and allows collaboration options to work in a team with ease.

b. Object-Oriented Analysis and Design (OOAD) is the design methodology used for this project. This approach is tailored for developing systems which are object oriented, thus aligning with the development of machine learning models. It also encourages the creation of modular and reusable components, which is advantageous. It also ensures the development aspect of this project is well-structured and efficient.

c. Both evaluation metrics and benchmarking have been used in this project as the evaluation methodology. A set of evaluation metrics will be used to quantitatively assess the performance of this machine learning model. The metrics will include classification metrics to evaluate the presence of osteoarthritis and regression metrics to evaluate the severity of osteoarthritis. Treatment recommendation metrics will ensure enhancing the recommendations. Benchmarking will also be used to contextualize the model's performance and validate its real-world capability. Here, this model will be compared to the existing models and provide valuable insights for the betterment of this project.

3.4 Project management methodology

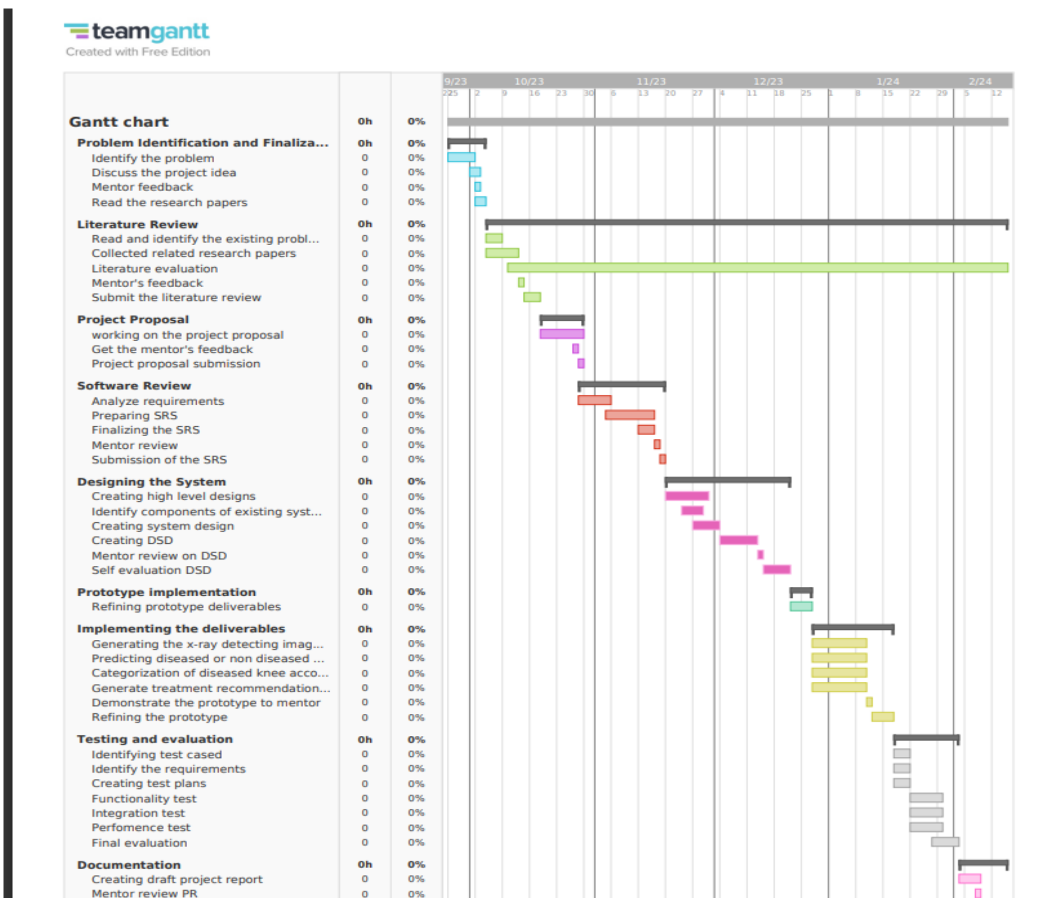

Figure 3 Gantt chart

Table 7 Deliverables

Deliverable	Date
Semester 01	
Submission of literature review	Week 3
Submission of the Project Proposal to the supervisor	Week 4
Submission of the Project Proposal (Final PP)	Week 5
Submission of the SRS to the supervisor	Week 8
Submission of the SRS (Final SRS)	Week 9
Semester 02	
Prototype implementation	Week 14
Testing and evaluation	Week 19
Documentation and final report submission	Week 23

3.5 Chapter summary

Detecting osteoarthritis in knee joints can be challenging. Therefore, following a well-defined set of methodologies is crucial to achieve a higher accuracy when detecting the disease. Additionally, it also helps in choosing the best approaches which will make the process of training the system more efficient and dependable. Furthermore, by adhering the selected methodologies, the research seeks to enhance the overall effectiveness of knee osteoarthritis detection and advance medical diagnostics.

Chapter 4: Software Requirements Specification

4.1 Chapter Overview

This chapter primarily focuses on obtaining user requirements and analyzing the data gathered. After doing a stakeholder analysis, more information regarding the stakeholders was obtained. The second topic covered was the best requirement elicitation technique and its outcomes. Drawing use case and context diagrams is another crucial step in demonstrating how internal and external parties interact. Finally, both functional and nonfunctional requirements are outlined and arranged in accordance with the system's priority.

4.2 Rich Picture

This project uses Convolutional Neural Network for x-ray image recognition. X-rays will be read into the system and the necessary features will be extracted to confirm if the x-ray is of a knee joint. Once the knee joint is confirmed, a binary classification model will be used to check if the knee joint has osteoarthritis or is normal. The system will use a Decision Tree model to detect the severity of the osteoarthritis and then use a random forest model to recommend suitable non-medical treatment based on the severity.

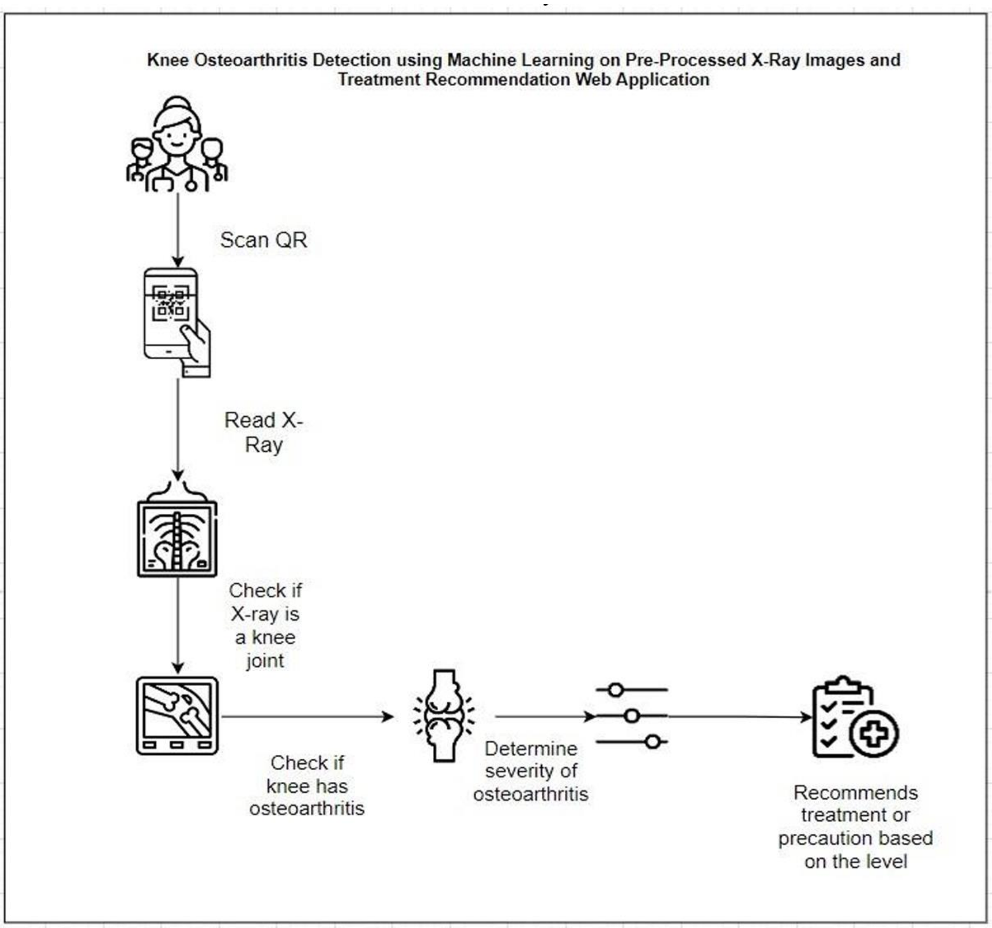

Figure 4 Rich Picture

4.3 Stakeholder Analysis

4.3.1 Onion Model

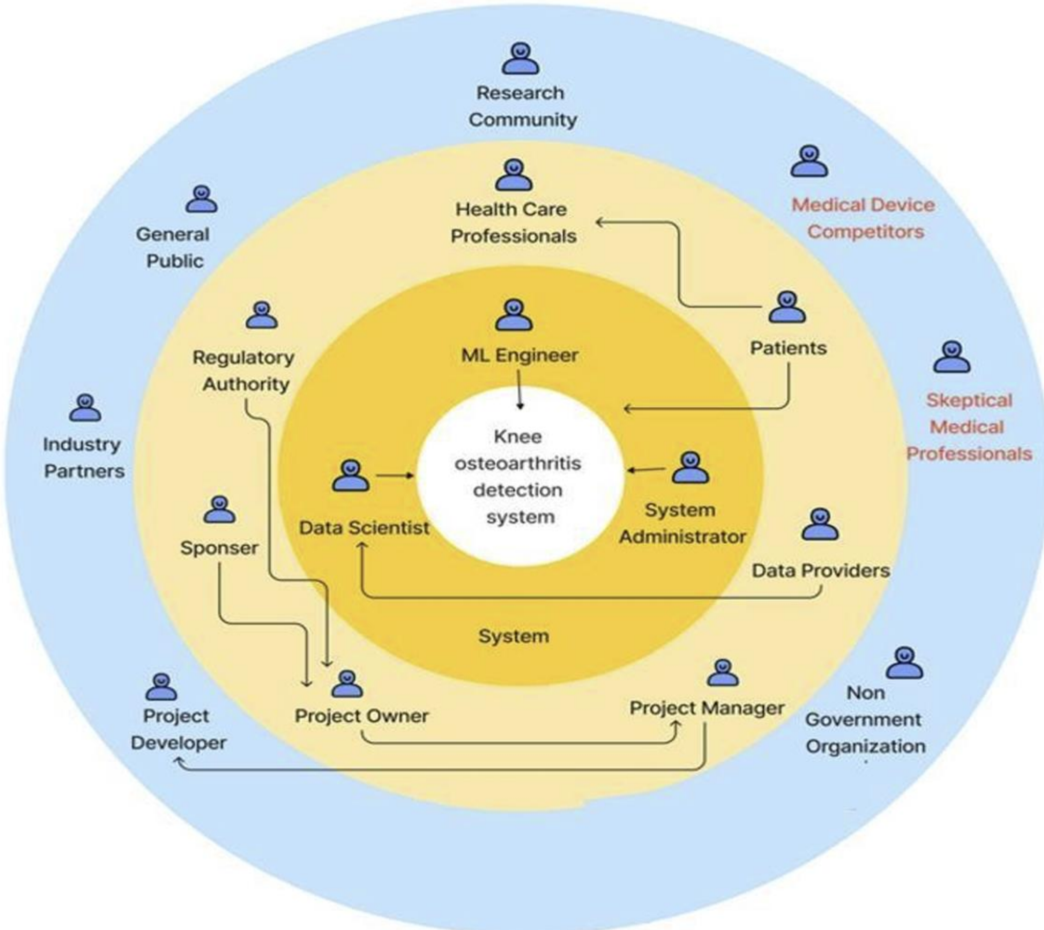

Figure 5 Onion Model

4.3.2 Stakeholder Viewpoints

Table 8 Stakeholder Viewpoints

Stakeholder	Role	Benefits
Data Scientist, ML Engineer	Data Analysis and Interpretation	Both roles involve tackling complex problems using data. They benefit from the challenge of finding innovative solutions.

System Admin	Infrastructure Management	Efficient management by system admins ensures that the systems are reliable and consistently available
Data providers	Data supply	Contribution to the advancement of medical knowledge and innovation, leading to improved healthcare.
Patients	Data Provider	Provide X-rays for training and testing to enhance the system.
Health Care Professionals	End User	Enhanced system for diagnosis, treatment, and patient care, improving efficiency and accuracy in healthcare delivery.
Regulatory Authority	Safety oversight	Assurance of safe and effective use of medical data.
Sponsor	Financial Support	Investments for advancements in healthcare
Project Owner	Project Oversight	Successful project outcomes and contribution to healthcare innovation

Project Manager	Resource Management	Successful and timely completion of the project within the allocated resources and meeting the defined objectives.
Project Developer	Project planning and implementing	Contribution to innovation and professional growth.
Industry Partners	Collaborative Support	Shared expertise, resources, and potentially marketable products or services.
General Public	End Beneficiaries	Access to better healthcare services, improved quality of life, and overall health improvements.
Research Community	Expertise Contribution	Contribution to scientific knowledge and access to new findings.
Medical Device Competitors	Market Competition	Competing products can drive innovation, potentially leading to improved offerings.
Skeptical Medical Professionals	Risk Assessment	Ensuring the safety and efficiency of new technologies, preventing potential risks for patients.
Non-Government Organization	Service Provision	Contribution to social well-being and the promotion of ethical healthcare practices.

4.4 Selection of Requirement Elicitation Techniques

Requirement elicitation is the process of obtaining information from stakeholders to recognize their needs and expectations. The following examines the methods used and the advantages/disadvantages of them.

Interviews provide an opportunity to gain a comprehensive understanding of stakeholders' viewpoints, expectations, preferences, and acquire more detailed information pertaining to the system. Some advantages of interviews are that interviewers can ask questions and seek clarification in real time, stakeholder engagement, and early problem identification. This method also has its downsides such as the time taken and effort to schedule and conduct the interview, limited number of participants and limited number of questions due to time constraints.

Followed Requirement Gathering Methods

Surveys and Questionnaires without Clinical Input

Since this project focuses on knee osteoarthritis surveys and questionnaires without clinical input from healthcare professionals may result in overlooking critical clinical aspects. Finding the target audience would also be a challenge as it needs to be specified by medical professionals. Therefore, making this approach unsuitable for the project.

Document analysis

This method allows the chance of having outdated or incomplete information. Limited stakeholder involvement also makes this method not suitable for this project.

Brainstorming without Technical Constraints

Unconstrained brainstorming may generate ideas that are not feasible or practical from a technical or medical standpoint therefore making this method unsuitable for the project.

Optimal option

The optimal option for this project is Interviews since more information will be gathered within a limited time with a biased opinion therefore allowing to explore multiple perspectives into the problem.

4.5 Discussion of Results

Two doctors were identified as stakeholders and were interviewed to gather their opinions and views on this project. One of the doctors is a general practitioner and professor, the other doctor is an oncologist.

Question 1: What do you think about this project?

Aim of this question: To get an idea about if doctors will use this system and if they are willing to rely on Artificial Intelligence for diagnosing.

Summary of findings: The doctors were highly impressed with this project. They highlighted their trust in artificial intelligence to computerize the diagnosing of knee osteoarthritis. Several systems to diagnose diseases from MRI scans are currently used, thus proving that they are open to change and trust new technologies.

Conclusion: Even though many health care professionals are still skeptical about using technology in health care practice, there are several practitioners willing to accept change and let AI be advantageous for them.

Question 2: How easy is it to determine the level of KOA?

Aim of this question: To determine if automating the process of determining the level of osteoarthritis is necessary.

Summary of findings: The process of determining the grade of Knee Osteoarthritis is challenging and not always accurate. Treating knee osteoarthritis according to the level of it, can increase the rate of recovery drastically.

Conclusion: Determining the severity of knee osteoarthritis is a crucial step to provide more suitable and effective treatment plans.

Question 3: What do you think about determining the level of Knee Osteoarthritis using the KL grading system?

Aim of this question: To determine the best method to classify knee osteoarthritis.

Summary of findings: From the several methods of determining the severity of knee osteoarthritis, the most used is the KL grading system.

Conclusion: The best method of classification is the KL grading system since it is widely used, and most doctors are familiar and comfortable with this grading system.

Question 4: What do you think about automating this process?

Aim of this question: To determine the complexity of manually diagnosing the severity by looking at an x-ray.

Summary of findings: A knee osteoarthritis specialist will be able to determine the severity much more easily than the other doctors. However, for doctors not specialized in this field, it can be quite challenging to determine the severity at first look.

Conclusion: Adding a feature to classify the knee with osteoarthritis according to the severity will help doctors who are not specialized in the field of osteoarthritis.

Question 5: Who will benefit most from our project (what kind of doctors)?

Aim of this question: To determine the user.

Summary of findings: Allowing doctors not specialized in the field of osteoarthritis to be the main user of this system is more suitable since not everyone with knee pain first consults the osteoarthritis specialist. Usually, patients visit general practitioners who then recommend specialists after diagnosing the disease.

Conclusion: This project will target general practitioners instead of the Knee Osteoarthritis specialists as it will be more useful for the earlier.

Question 6: What do you think about treatment recommendations based on severity?

Aim of this question: Recommending treatment is a feature of this system. This question is to determine if this feature is useful.

Summary of findings: Doctors were skeptical about recommending medical treatment to patients as there are many factors to consider before recommending medicines to patients. However, they encourage recommending non-medical treatment such as diet plans and exercise plans to improve knee health based on the level of osteoarthritis.

Conclusion: Recommending non-medical treatment plans is a useful function.

Question 7: Do you have any recommendations to enhance this project?

Aim of this question: To enhance this project to fit the requirements of the doctors.

Summary of findings: One main factor that concerned doctors was how the x-ray image will be uploaded into the system. The image needs to be of high quality, and they will always need other devices with them, to be able to take pictures of the x-ray and upload it into the system.

Conclusion: A way of directly sending the soft copy of the x-ray to the doctor needs to be figured out.

4.6 Summary of Findings

Although there are several doctors still skeptical about using artificial intelligence to diagnose diseases, there were doctors ready to adapt to new changes as well. Through the interviews conducted it was prominent that this system will be put to good use, and doctors do trust the accuracy of such systems. Futherthemore, it was noted during these interviews that this system will mostly benefit doctors who are not specialized in the field of osteoarthritis, therefore our target users will be the general practitioners. The main features proposed in this system, such as the x-ray reading, severity classification and non-medical treatment recommendation aligned with the requirements of the doctors, however, they raised a concern about uploading the image to the system. Thus, it has been decided to use a QR code system to make the process easier. The radiologist scans the x-ray and issues a QR code which will contain the x-ray image of the patient. The doctor will scan this QR code to extract the x-ray, instead of having to take a picture of the x-ray hardcopy and upload to the system.

4.7 Context Diagram

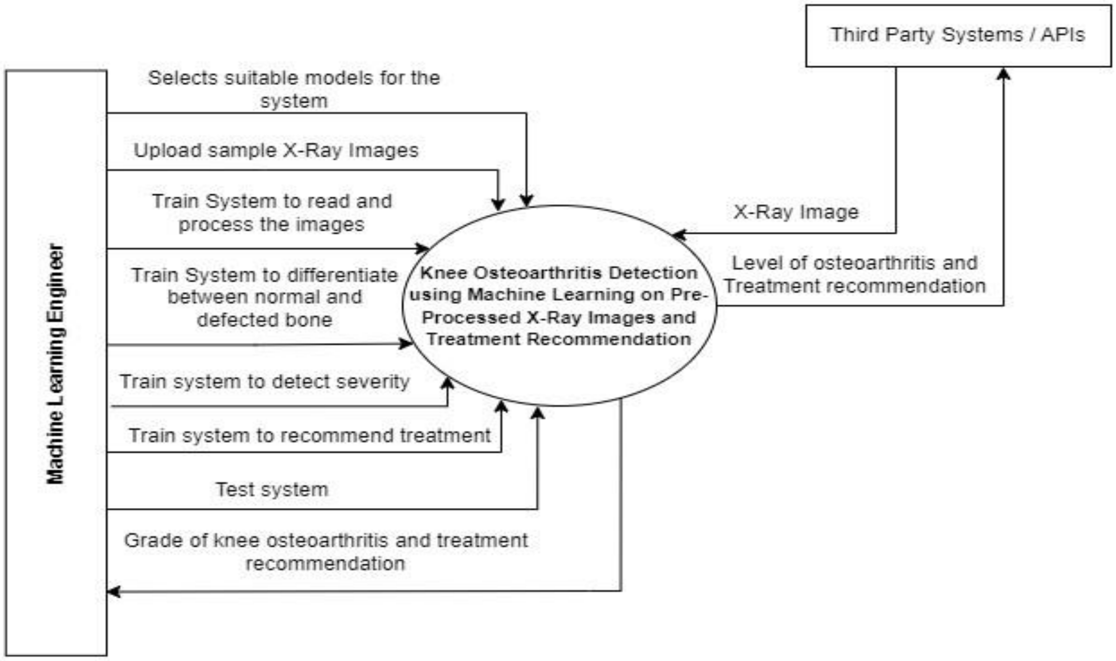

Figure 6 Context Diagram

4.8 Use Case Diagram

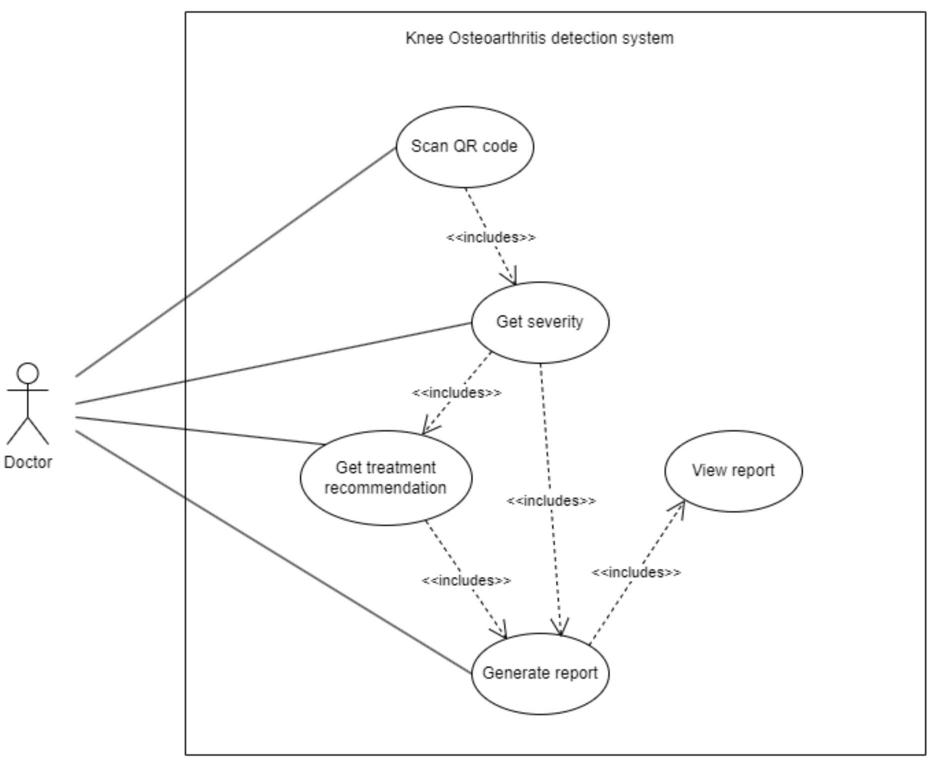

Figure 7 Use Case Diagram

Use case Descriptions

Table 9 Use Case Description

Use case ID	UC01
Use Case Name	Scan QR code
Description	Scan QR code to obtain X-ray
Participating Actors	Doctor
Pre-Conditions	Must be a QR code
Main Flow	1. User inputs QR code. 2. QR code is scanned and X-ray is taken. 3. X-ray is processed. 4. It is graded according to the severity and a report is generated.
Alternative Flow	-
Exceptional Flows	E1. User input is not a QR code. ● UC01 is repeated.

Use case ID	UC02
Use Case Name	Get severity
Description	Gives the level of severity of Osteoarthritis
Participating Actors	Doctor
Pre-Conditions	X-ray is obtained from the QR code
Main Flow	1. The severity is graded. 2. Treatment recommendations given. 3. Report is generated.
Alternative Flow	-

Exceptional Flows	E1. Model fails/Doesn't grade • Redo UC02

Use case ID	UC03
Use Case Name	Get treatment recommendation
Description	Gives the treatments according to the severity
Participating Actors	Doctor
Pre-Conditions	Severity of Osteoarthritis is classified
Main Flow	1. Give treatment recommendations. 2. Generate report.
Alternative Flow	-
Exceptional Flows	E1. Model fails • Redo UC03

Use case ID	UC04
Use Case Name	Generate report
Description	Make a report for the user to view.
Participating Actors	Doctor
Pre-Conditions	Severity is graded, and treatment recommendations given.
Main Flow	1. Report is generated 2. Report is viewed.

Alternative Flow	-
Exceptional Flows	E1. Report is not generated. • Use case fails.

4.9 Functional Requirements

Table 10 Functional Requirements

	Requirement and Description	Priority
FR01	QR code reading and image processing: X-ray images will be uploaded into the system by scanning a QR code which contains the x-ray images, these images need to be processed.	Critical
FR02	Extract Features from X-rays: To determine if the x-ray is of a knee joint or not, the features need to be extracted and examined.	Critical
FR03	Osteoarthritis detection: The system must first check if the knee joint is normal or is affected with osteoarthritis.	Critical
FR04	Severity Assessment: The severity of osteoarthritis must be detected to provide a tailored treatment recommendation system.	Critical
FR05	Treatment Recommendation: Treatment plans need to be recommended according to the severity of the disease.	Critical

FR06	User Interface: User friendly interface to ease the process of uploading x-rays and reading the processed outcome.	Important
FR07	Continuous learning: The system must learn whenever tested with new datasets.	Non- Important

4.10 Non-Functional Requirements

Accuracy

Increase the precision of a deep CNN's detection of knee osteoarthritis by using transfer learning to large-scale medical imaging datasets. Optimize settings and utilize data enhancement to achieve increased specificity and sensitivity.

Performance

Improve knee osteoarthritis detection by using advanced features and combining multiple learning models. Regularly update the system based on feedback from healthcare experts for ongoing accuracy.

Security

The system should be secured to avoid unauthorized access as the process includes interaction with sensitive personal data of the patient such as x-rays.

Usability

Creating a user-friendly user interface, giving clear directions, and including features that are easy for users to utilize. Iterative design approaches can benefit from incorporating user feedback to improve overall user experience and improve workflow.

Compliance with Medical Standards

Enhance the system by working with healthcare providers, conducting routine reviews of medical literature, and putting in place a dynamic updating mechanism. To make sure the system's recommendations are in line with the most recent research and medical practices, create a feedback loop with users and interface with electronic health records for real-time patient data.

Table 11 Functional Requirements

	Requirement and Description	Specification	Priority
NFR01	Identifying the knee bone and Osteoarthritis detection should have high level of accuracy	Accuracy	Important
NFR02	Severity assessment of osteoarthritis should be precise	Accuracy	Important
NFR03	Fast response time when analyzing and processing x ray images	Performance	Important
NFR04	X ray images and patient data should store and transmitted securely according to privacy regulations	Security	Important
NFR05	User interface should be intuitive and easy to use	Usability	Non-Important
NFR06	Machine learning models used for osteoarthritis detection meets medical standards.	Compliance with Medical Standards	Important
NFR07	System needs to be regularly updated according to the latest research on the medical field.	Compliance with Medical Standards	Non important
NFR08	Be able to update and maintain the system easily	Maintainability	Non- important

4.11 Chapter Summary

This chapter explores the methodologies of requirement elicitation, highlighting the need of comprehending stakeholder viewpoints using tools such as the onion model. It examines how to choose the best elicitation strategies, talks about the outcomes, and presents the data in a tabular format with obvious sourcing. The project scope is clearly illustrated by diagrams such as the use case and context diagrams. Clearly focused functional needs were given priority, whereas non-functional requirements focused on system quality.

Chapter 5: Social, Legal, Ethical and Professional Issues

5.1 Chapter Overview

SLEP analysis is a tool that allows to evaluate several factors that can affect the product proposed in this thesis. In order to maximize the product's potential, the social, legal, ethical and professional issues are taken into consideration and strategies too mitigate them are discussed. This allows to ease the process of decision making, strategic planning and problem solving.

5.2 SLEP issues and Mitigation

5.2.1 Social Issues

- Patients may be concerned about their private health information used in this system being shared or stolen. To mitigate this, patient information must be protected with high priority.
- Doctors might need to learn new things to use the detection system, thus the system needs to be designed in a simple and easy to learn manner.
- This system is proposed to simplify the detection of knee osteoarthritis and provide basic non-medical treatments. However, these treatments are not enough to completely cure the disease, therefore if the doctor using this system is not a knee osteoarthritis expert, it has been advised to refer the patient to a specialist for better treatment.
- Introducing new technology might change how some medical jobs are conducted. The aim should be to assist medical experts and not replace them.

5.2.2 Legal Issues

- The data used for the project were taken from Kaggle and used without any additions, transformations or changes to the original dataset. The creator of the dataset has been credited to comply under Creative Commons licenses.
- Content of the literature review taken from research papers and articles were obtained from official sources and referenced to the respective authors.
- The model's limitations, risks and assumptions, along with protocols for handling errors will be clearly documented.
- The terms and conditions of software tools and libraries used that are licensed under General Public License are complied with.

5.2.3 Ethical Issues

- It is crucial to guarantee the security, privacy, and confidentiality of patient data that the system collects and uses. Additionally, to guard against breaches, illegal access, and patient information leakage.
- Patients need to be properly informed about the use of their data in such a system. They have to be aware of the advantages and disadvantages that could arise, as well as how their data will be managed and safeguarded.
- The system can help medical professionals to diagnose patients' illness and prescribe treatments, but it should not take the place of expert judgement. When it comes to patient

care, doctors or the clinicians should have the last say. The system should only be a tool to help them make the decisions.

- All hospitals should have equal access to the system and its suggestions, regardless of their financial situation or geographic location. This would guarantee that all patients receive the same caliber of care and recommendations.

5.2.4 Professional Issues

- Emails were used to communicate with mentor and schedule weekly meetings.
- Mentor meetings were conducted regularly through Google meet and physically, to discuss team progress and obtain feedback.
- Interviews were conducted with stakeholders to get better insights.
- WhatsApp was used to communicate among group members.
- Git and GitHub was used for version controlling and collaboration.

5.3 Chapter Summary

This chapter explores the various social, legal, ethical and professional issues that were encountered while developing the product. Finding solutions to mitigate these issues are mandatory in order to reduce the potential risks and challenges that may arise during the deployment stage of the product.

Chapter 6: System Architecture and Design

6.1 Chapter Overview

In this chapter, a thorough plan and visual representation of the system, featuring design paradigms, component diagrams, class diagrams, sequence diagrams, UI/UX designs, and process flowcharts can be found. These illustrations are instrumental in comprehending the system's architecture and the interactions among its components.

6.2 Design Goals

This table shows the standards that should be kept when making this system such as compatibility, security and more.

Table 12 Design Goals

Design Goals	Description
Compatibility	Must be compatible with various web browsers to maximize accessibility.
Responsiveness	Website should be made to be responsive especially on desktops to maximize user experience.
Security	Only medical professionals inside the specific medical center should be able to access, which should be done by making security measures such as encryption and secure authentication methods.
Speed	The models should provide fast inference speed to provide timely feedback to users.
UI/UX	Website interface should be user friendly and should use accessibility evaluations to improve user experience.

6.3 System Architecture Design

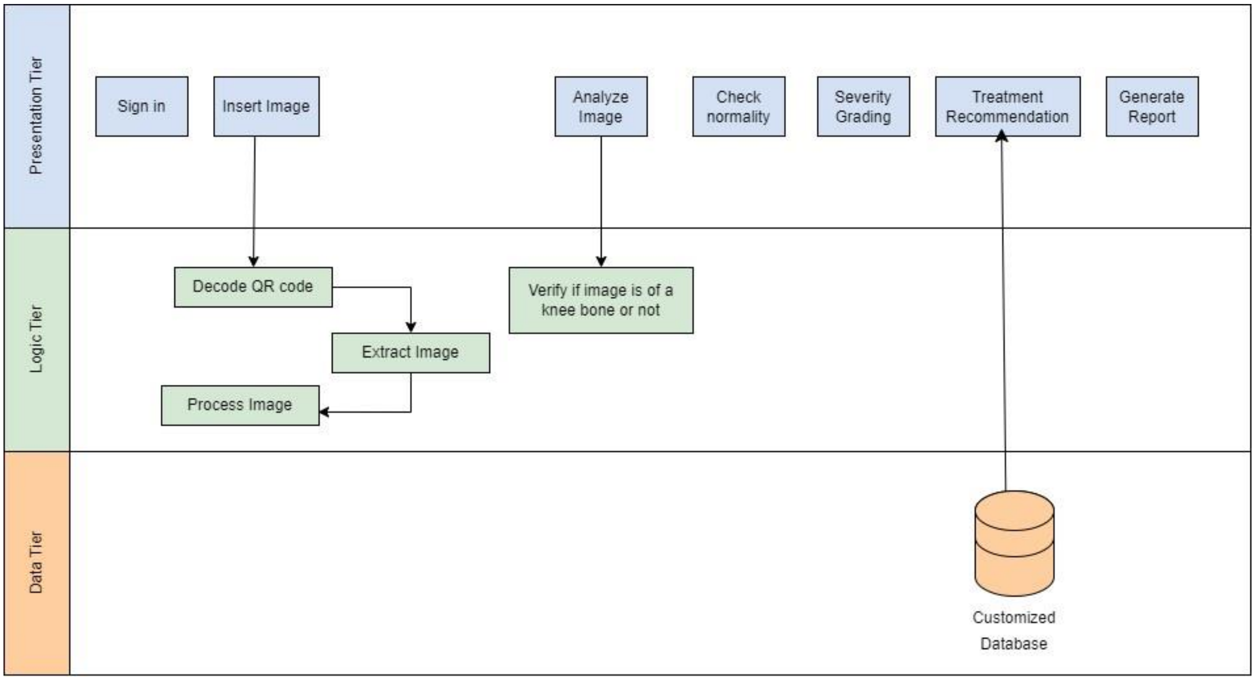

Figure 8 System Architecture Design

The above diagram is the architecture design of the system. This has three layers: presentation tier, logic tier and data tier.

1. Presentation Tier:
    - Sign in: this allows users from registered hospitals access the facilities the system provides.
    - Insert Image: allows users to insert the x-ray for analyzing. The user can either directly upload an x-ray image or the QR code that contains the x-ray.
    - Analyze Image: this analyzes the image to check if the image is of a knee bone or not.
    - Check normality: checks if the knee bone is normal or not.
    - Severity Grading: predicts the severity of the diseased knee.
    - Treatment Recommendation: generates treatment recommendation based on the predicted severity grading.
    - Generate report: all of the above will be displayed for the user to view as a report after all processing takes place.
2. Logic Tier
    - Decode QR code: if the user inserts a QR code of the x-ray, it must be decoded first.
    - Extract Image: the image is extracted from the decoded QR code for processing.

- Process Image: the relevant processing is done to the image such as resizing, normalizing and dimension expansion to make a common format for analyzing.

3. Data Tier
   - Customized Database: a database was created to store the treatment recommendations for each severity level of knee osteoarthritis.

6.4 System Design

There are two types of system design methodologies, which are:

1. SSADM: Structured System Analysis and Design Methodology
2. OOAD: Object-Oriented Analysis and Design

6.4.1 Choice of Design Paradigm

Object-Oriented Analysis and Design (OOAD) is the design methodology used in this project as it is an approach made for developing projects which are object oriented. This method ensures a well-structured and efficient development workflow, showing clearly the responsibilities among components, which ensures maintainability and scalability. This also encourages encapsulation of functionality within objects, resulting in enhancing code readability and reducing complexity.

Since this project works with an object-oriented nature, Structured Systems Analysis and Design Methodology (SSADM) cannot be considered as a suitable design paradigm. Object-Oriented Analysis and Design (OOAD) allows modularity and reusability, which is considered crucial when developing machine learning models as most of the time algorithms and other techniques need to be used in several other places. Furthermore, considering the dynamic nature of machine learning systems it can be concluded that the most suitable design methodology is the Object-Oriented Analysis and Design (OOAD) because of its flexibility and adaptability to handle the dynamic nature of machine learning projects.

6.4.2 Component Diagram

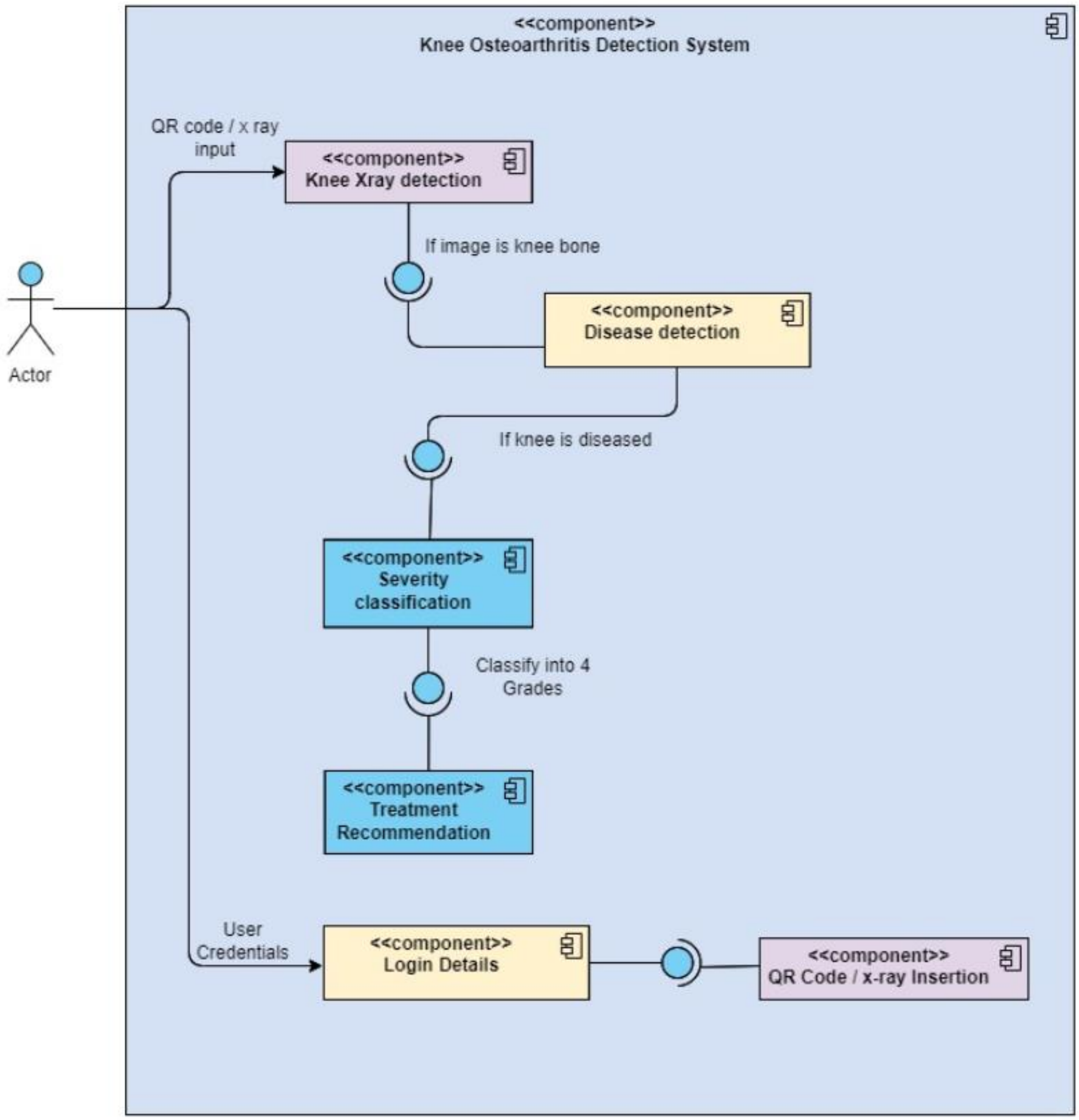

Figure 9 Component Diagram

6.4.3 Class Diagram

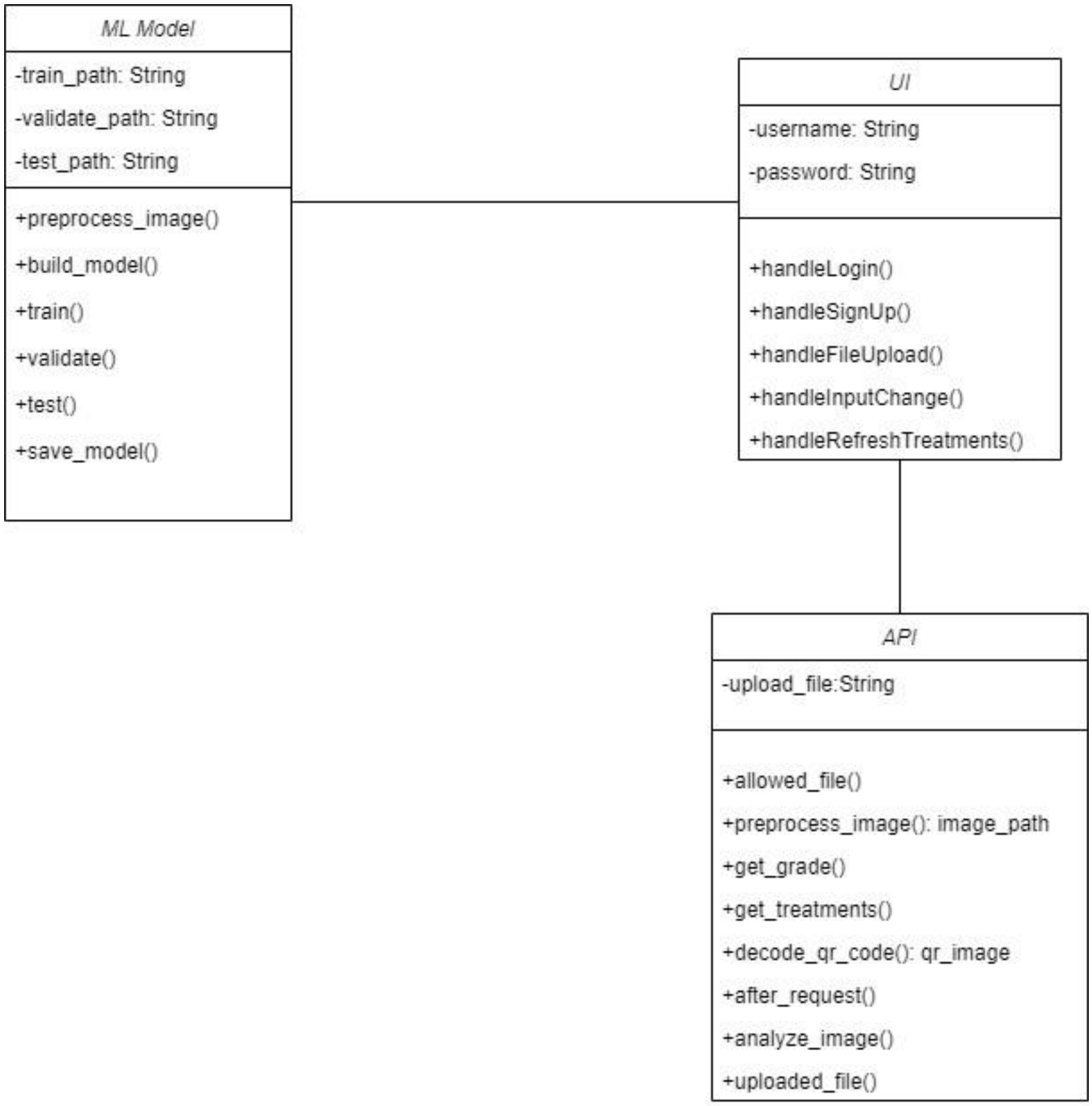

Figure 10 Class Diagram

6.4.4 Sequence Diagram

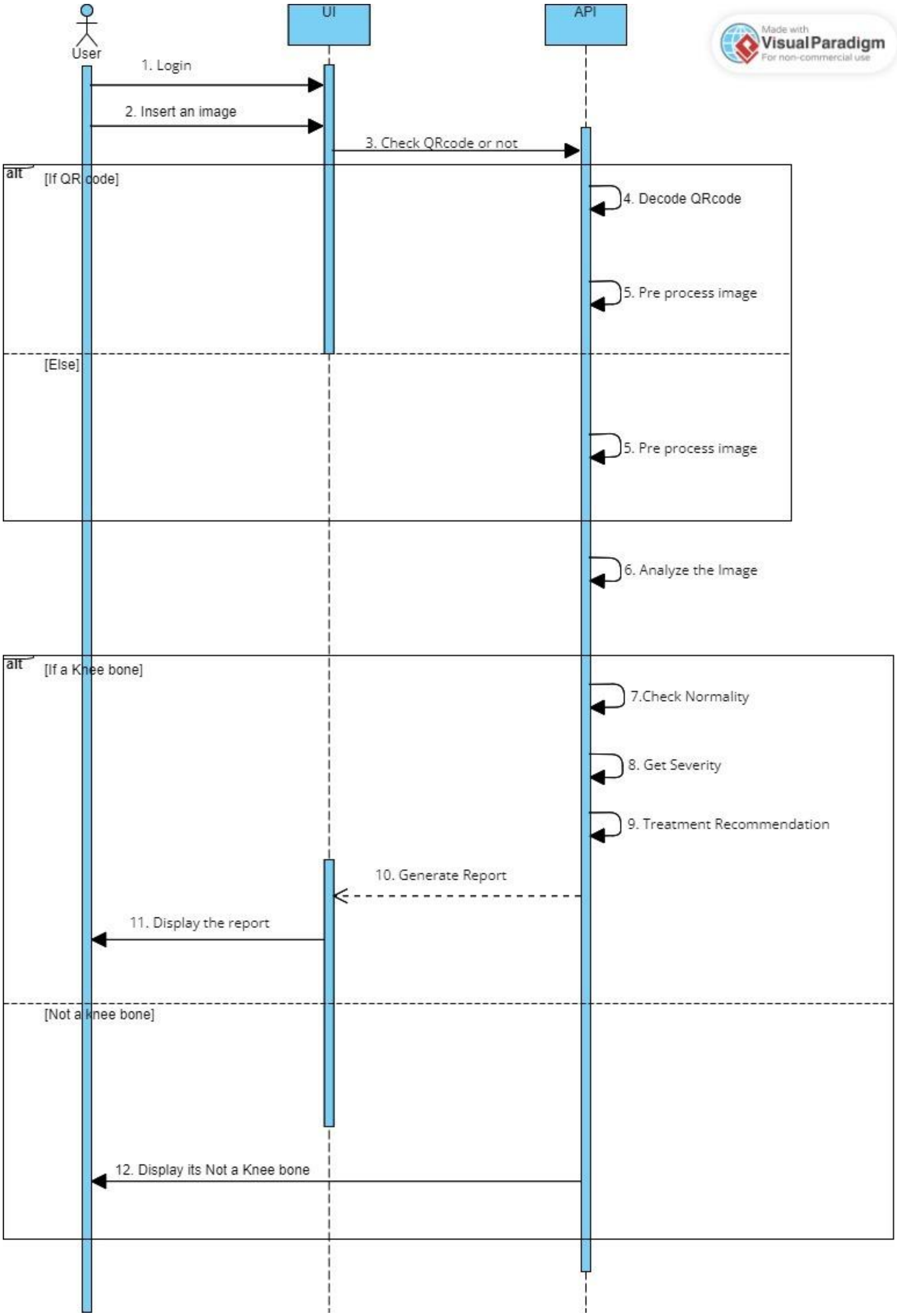

Figure 11 Sequence Diagram

6.4.5 UI Design

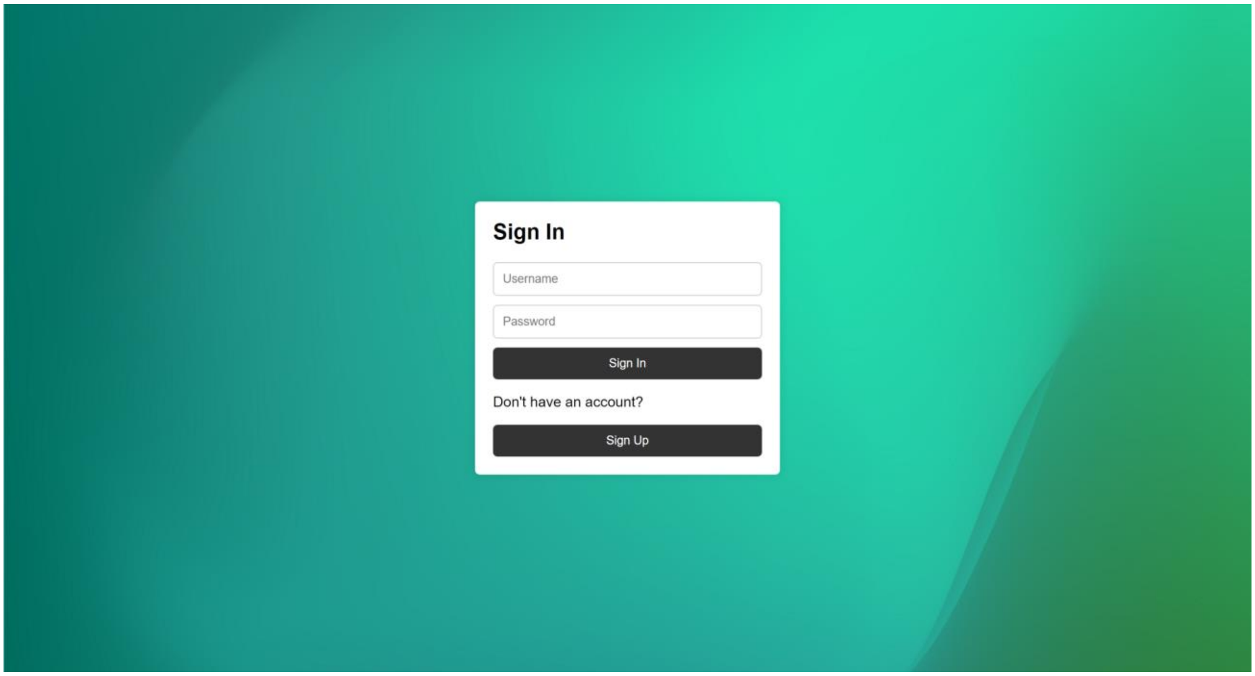

Figure 12Osteosense- Sign In page

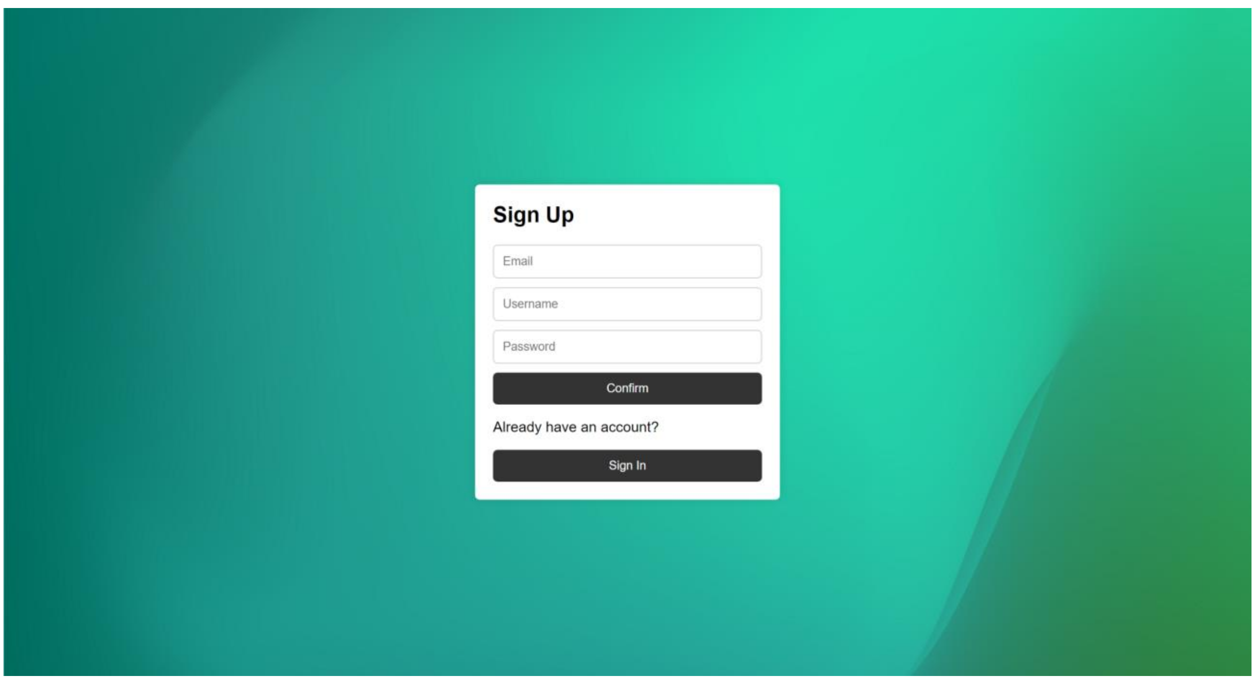

Figure 13 Osteosense Sign Up page

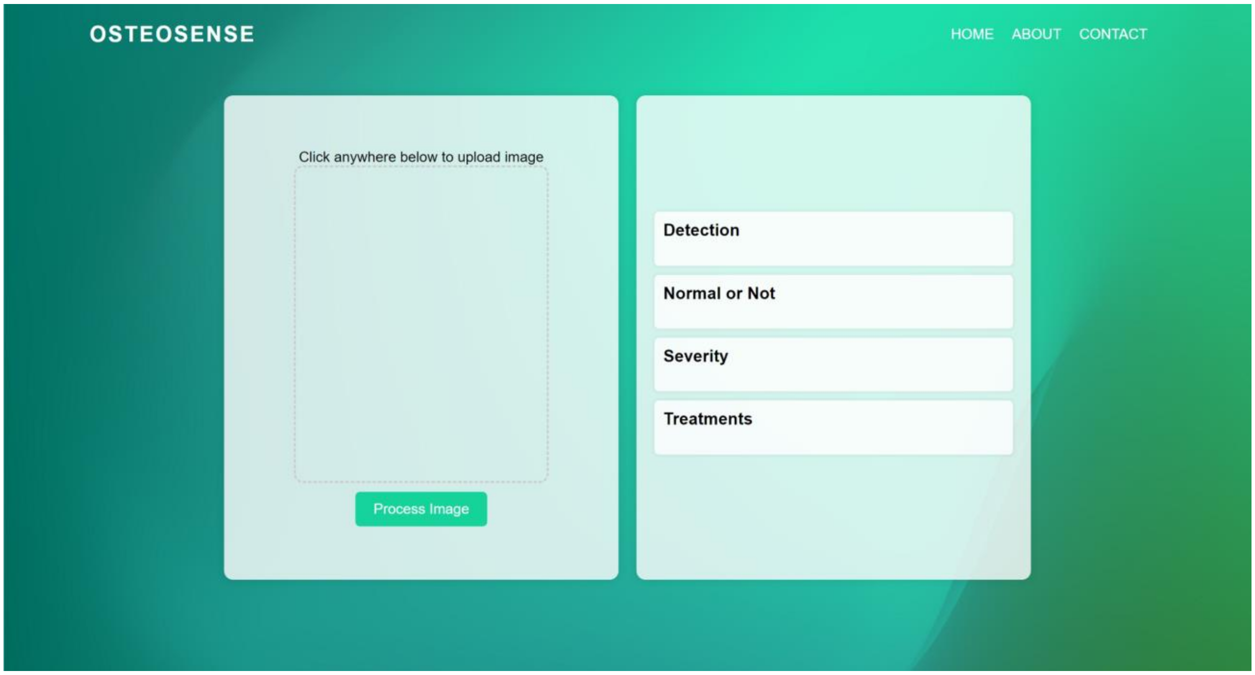

Figure 14 Osteosense Home page

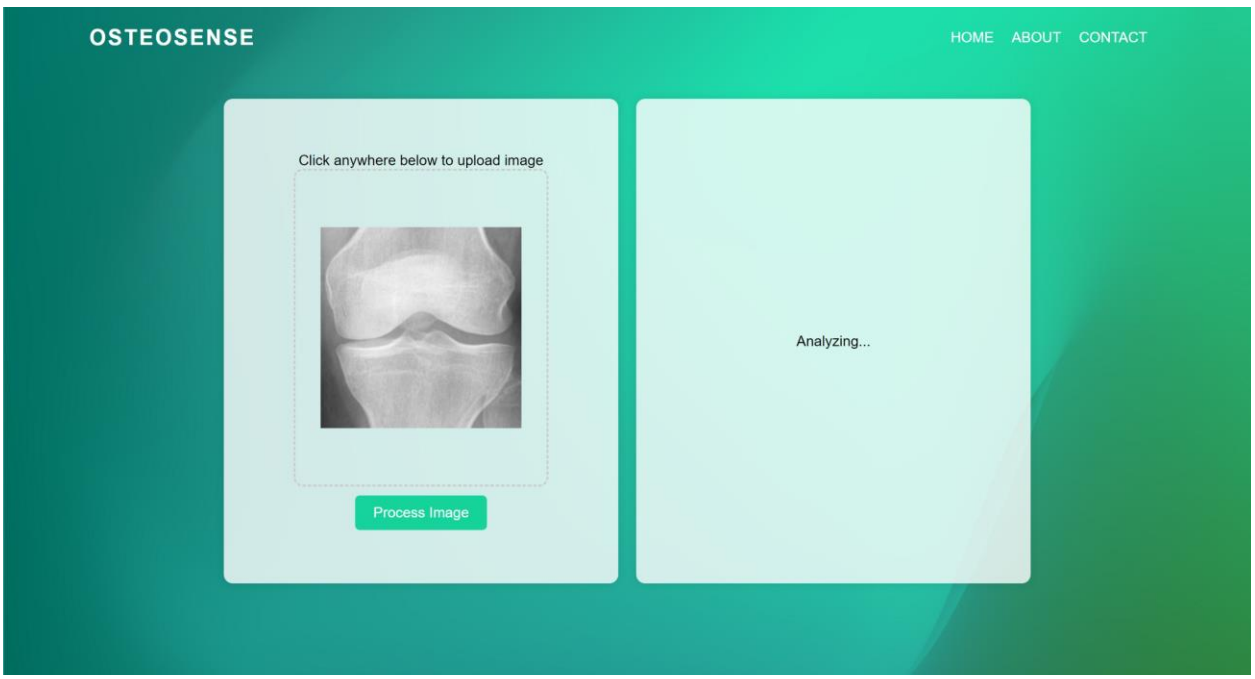

Figure 15 Osteosense- Image analyzing

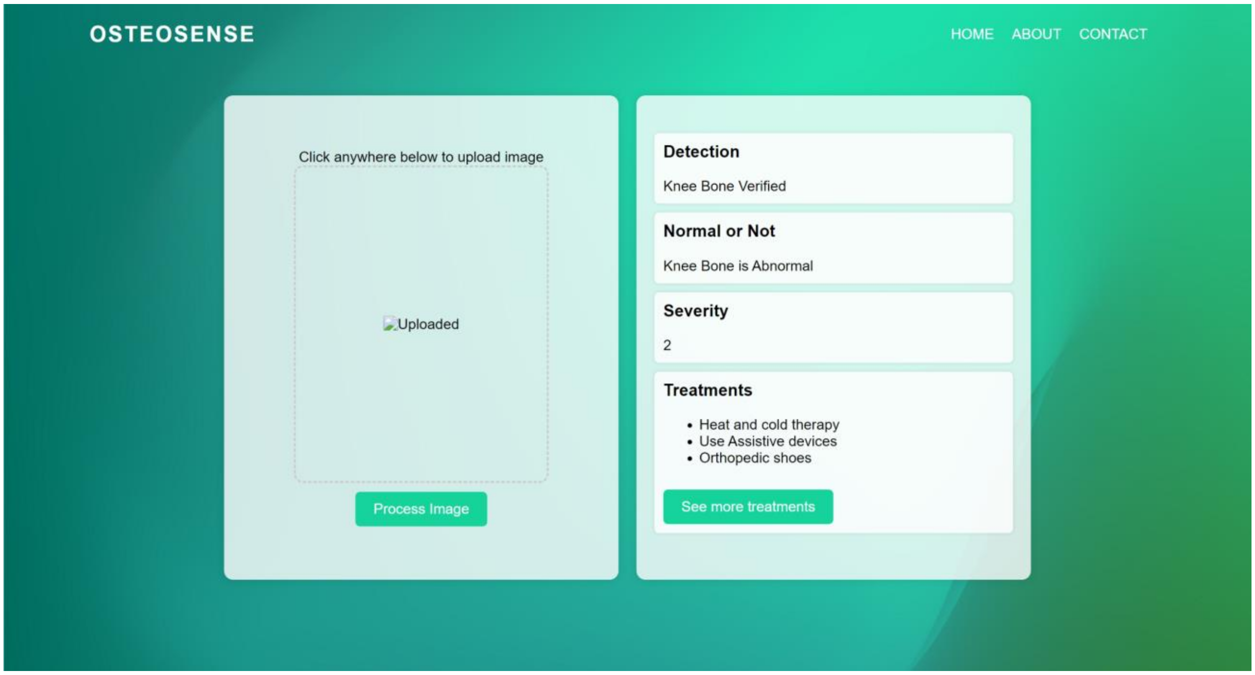

Figure 16 Osteosense- Report

6.4.6 Process Flow Chart

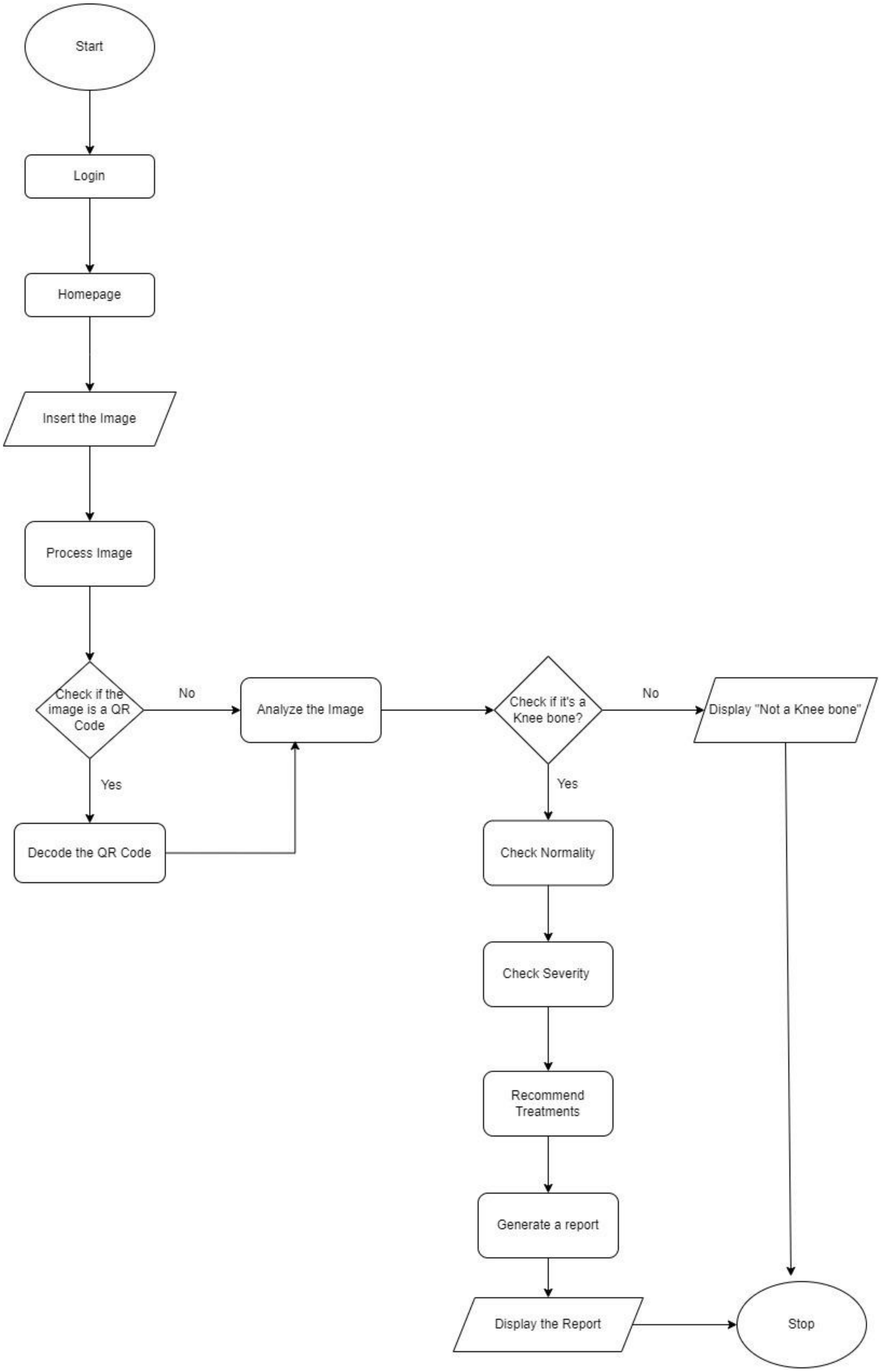

Figure 17 Process Flowchart

6.5 Chapter Summary

This chapter has outlined the diverse design elements of the system. The component diagram, class diagram, sequence diagram, and UI/UX designs showcased here offer a lucid representation of the system's intended design goals.

Chapter 7: Implementation

7.1 Chapter Overview

In this section we will look at the main tools and technologies used to build the knee osteoarthritis detection system. Below, we will evaluate the technology stack, the methodologies used for data selection, languages and frameworks used to create the project and the implementation in pseudocodes.

7.2 Technology Selection

7.2.1 Technology Stack

Multiple technologies have been used for the creation of the application. Mainly PyCharm application was used to build the machine learning models. For front end React along with HTML, CSS and JavaScript was used. For back-end Flask was used.

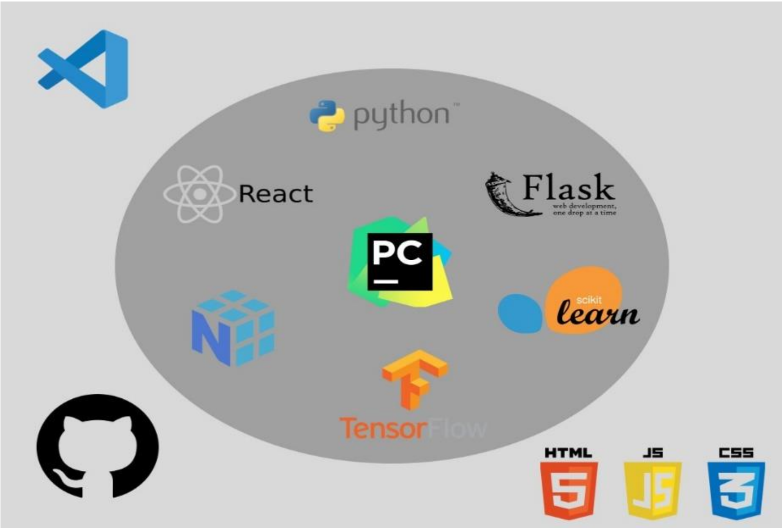

Figure 18 Technology Stack

7.2.2 Data Selection

The Selection of a dataset plays an important role in the whole process of building a machine learning model. This project required images of X-rays of knee joints, which was our primary objective to achieve when selecting a dataset. To achieve this, the Kaggle website was used which contained many datasets relating to this project.
Finally, the dataset was chosen by considering the reviews, number of downloads and the usability rating given by the website which was the best among the many other datasets.

Table 13 Data Selection

Domain	Dataset	Description

Knee Osteoarthritis Detection Using Machine Learning on Pre-Processed X-ray Images	Knee Osteoarthritis Dataset with Severity Grading	The dataset includes knee x-ray images, going from grade 0 to 4 according to Kellgren-Lawrence (KL) scale, indicating the severity of osteoarthritis. The dataset, sourced from the Osteoarthritis Initiative (OAI), is publicly available and was compiled by Pingjun Chen in 2018. Link to dataset: https://www.kaggle.com/datasets/shashwatwork/knee-osteoarthritis-dataset-with-severity/data
Treatment Recommendation	Custom dataset	The dataset includes several treatment options for each grade according to the Kellgren-Lawrence (KL) scale.

7.2.3 Selection of Development Framework

Frontend

React was utilized to create a web application since it offers a streamlined approach to building modern interactive interfaces. It has a component-based architecture and declarative syntax which allows developers to break down complex UIs into reusable components, while assuring code reusability and maintainability. React has a rich ecosystem of libraries and tools which allow developers to create dynamic and responsive web applications, which makes it a preferred choice for front end development.

Backend

Python with Flask was used for the backend development if the web application since it provides a robust and efficient foundation for handling data processing, routing and API development. Python is widely known for its simplicity, readability and extensive library ecosystem which makes it an ideal choice for backend development. Flask offers a lightweight and flexible framework for building web applications as it allows developers to set up routes and manage HTTP requests, ensuring smooth communication between the frontend and backend components.

7.2.4 Programming Languages

Python is the primary programming language used in the implementation of this web application for knee osteoarthritis detection and treatment recommendation system alongside JavaScript, HTML and CSS.

Python serves as the backbone of the project, enabling the implementation of ML components that are used in this project through its versatile and extensive ecosystem of libraries and frameworks. TensorFlow and Keras are used to design and implement CNN models while VGG16 architecture enhances the feature extraction for improved classification accuracy. Also, the treatment recommendation system utilizes Scikit-learn and Random Forest algorithms for robust analysis and treatment suggestions.

Flask, a python web framework, is used to integrate our machine learning components into the web application, ensuring simplicity and efficiency throughout the development and deployment process. It facilitates the creation of RESTful APIs to handle user inputs, enabling efficient processing of QR codes and X-ray images. React, a popular JavaScript library for building user interfaces, is also used to develop the frontend, using its component-based architecture for creating interactive and responsive interfaces.

7.2.5 Libraries

Table 14 Libraries

Libraries	Version
TensorFlow	2.15.0
Keras	2.15.0
Numpy	1.26.3
Sklearn	1.4.1.post1
Pandas	2.2.1
Flask	3.0.2
Matplotlib	3.8.2
Cv2	4.9.0
joblib	1.3.2

7.2.6 IDE

PyCharm was used as the primary IDE for developing this system. It has a comprehensive set of features that facilitates efficient coding, debugging and version controlling throughout the development process. Google Collab was also used to enhance the efficiency of model training within the system. It facilitated training machine learning models well since it has access to powerful GPU resources.

7.2.7 Summary of Technology Selection

Table 15 Summary of Technology Selection

Component	Technology/Tool	Version
Programming Language	Python	3.11.0
UI Frameworks	React	18.2.0
IDE	PyCharm	2023.2.5
	Google Collab	

7.3 Implementation of Core Functionalities

Component 1: Knee Bone checker

This component focuses on checking if the input image is of a knee bone or not.

```
BEGIN
        define class to identify knee bone:
                define function to initialize variables:
                        input_shape ← dimensions of input images
                        model ← models.Sequential()

                define function to build_model:
                        Construct sequential model
                                SET convolutional layers, max pooling layers
                                Flatten the output
                                Dense layers with ReLU activation
                                Optional dropout layer for regularization
                                Output layer for binary classification with sigmoid activation
                                Compile model with Adam optimizer and binary crossentropy loss

                define function to set_up_data_generators:
                        SET train_datagen and validation_datagen for real-time data augmentation on images and rescale pixel values to [0, 1]
                        SET train_generator and validation_generator to flow from directories with binary classification and specified classes
                        Configure generators with batch size, target size, and other augmentation parameters

                define function to train_model:
                        Train the model with train_generator
                        Set number of epochs
```

Validate the model on validation_generator
Display validation loss and accuracy

define a function to save_the_model:
Save the trained model to a file

Initialize class
END

Component 2: Knee Normality Checker
This component checks if the knee bone is normal or not.
BEGIN
define class to check normality of knee:
define function to initialize variables:
train_path ← path to training data
validate_path ← path to validating data
test_path ← path to testing data
img_size ← size of image
batch_size ← size of batch
SET train_generator set to none
SET validate_generator set to none
SET test_generator to none
SET model to none

define function to preprocess image:
train_datagen ← ImageDataGenerator for real-time data augmentation on images and the image is rescaled to 1./255
validate_datagen ← ImageDataGenerator for real-time data augmentation on images and the image is rescaled to 1./255
test_datagen ← ImageDataGenerator for real-time data augmentation on images and the image is rescaled to 1./255

generate batches of augmented/normalized data from the image directories

define function to build_model:
Load MobileNetV2
base_model ← MobileNetV2
SET base_model to non-trainable

Construct sequential model

```
                    SET base_model, global average pooling layer, dense layer, dropout
regularization, output layer with single neuron and sigmoid activation
                    Compile model

          define function to train model:
               SET epochs to 50
               train model

          define function to validate:
               GET train predictions from train_generator
               GET validate predictions from validate_generator
               GET train accuracy from train_generator
               GET validate accuracy from validate_generator
               GET classfication report for train
               GET classfication report for validate
               DISPLAY results

          define function for testing:
               GET test predictions from test_generator
               GET test accuracy from test_generator
               DISPLAY results

          define a function to save the mode:
               Save model
     Initialize class
END
```

Component 3: Osteoarthritis Severity Checker

This component detects the severity of the abnormal bone.

```
BEGIN
  DEFINE data_path ← path to training data
  DEFINE categories ← grade of severity
  DEFINE labels ← image labels

  CREATE dictionary ← {}
  FOR EACH category IN categories DO
    dictionary[category] ← labels[index_of(category)]

  SET image_size ← (width, height)
  INITIALIZE data_list ← []
  INITIALIZE label_list ← []
```

```
    FOR EACH category IN categories DO
        FOR EACH image IN category_folder(category) DO
            image ← READ(image)
            image ← CONVERT_TO_RGB(image)
            image ← RESIZE(image, image_size)
            APPEND image TO data_list
            APPEND corresponding_label TO label_list

    Normalize data_list
    Convert labels to categorical format

    LOAD pre-trained_model ← VGG16(top_layer=False)
    FREEZE layers in pre-trained_model

    RESIZE output of pre-trained_model
    ADD 1x1 Conv2D layer

    DEFINE custom_model ← Sequential()
    ADD convolutional_layers, max_pooling, flatten, dense_layers TO custom_model

    COMBINE pre-trained_model WITH custom_model

    COMPILE combined_model WITH loss_function, optimizer, metrics

    SPLIT data INTO training_data, testing_data

    TRAIN combined_model WITH training_data, specifying epochs, validation_split

    EVALUATE combined_model performance ON testing_data

    SAVE trained_model
END
```

Component 4: Treatment Recommendation

This component focuses on recommending treatment according to the detected severity.

```
BEGIN
    define class to recommend treatments:

        define function to initialize variables:
            train_path ← path to dataset
```

```
    SET dataframe to none
    SET X to none
    SET y to none
    SET X_encoded to none
    SET X_train to none
    SET X_test to none
    SET y_train to none
    SET y_test to none
    SET clf to none

  define function to load_dataset:
    Load dataset from the specified path

  define function to preprocess_data:
    Split the dataset into features (X) and target variable (y)
    Apply one-hot encoding to categorical variables

  define function to split_data:
    Split the encoded data into training and testing sets

  define function to build_model:
    Initialize the random forest classifier with specified parameters

  define function to train_model:
    Train the random forest classifier using the training data

  define function to evaluate_model:
    Make predictions on the test data
    Calculate accuracy of the model

  define function to save_model:
    Save the trained model to a file

  Initialize class
END
```

7.4 Chapter Summary

Mostly this chapter is focused on the technologies and tools used in the knee osteoarthritis detection system. One can gain a high-level understanding of the system's operation by examining the pseudocode implementation of key components such as knee bone identification, identify whether knee bone is diseased or not, severity classification and treatment recommendation. The following Chapter will analyze the testing phase of the system.

Chapter 8: Testing

8.1 Chapter Overview

This chapter addresses the system's performance evaluation and testing objectives. It discusses about assessing the four models in terms of accuracy, F1-score, precision, recall and a confusion matrix. This chapter covers functional testing, benchmarking, module and integration testing as well as non-functional testing components where load balancing, accuracy assessment and performance evaluation are performed.

8.2 Objectives and Goals of Testing

To ensure the deployed system satisfies all functionalities and requirements, to enhance the user experience is the primary objective of testing.
Testing is;

- To identify and mitigate potential vulnerabilities and errors that users might encounter, as it could compromise system's performance.
- To check whether the system performs as intended and aligns with user expectations. It ensures potential bugs and errors, which were not detected during development phase, are addressed promptly.
- To identify issues that could affect system's performance, stability, and security such as security vulnerabilities, and other possible failures.

By meeting these objectives, testing helps to lower the risks relates with software development and ensures that the end-product meets the expectations of the stakeholders.

8.3 Testing Criteria

Functionality Testing Criteria

- Image Recognition and QR Code Decoding:
  Verify that the system accurately recognizes and decodes QR codes, extracting the correct information.
- Knee bone detection:
  Test whether the system correctly identifies the input image as a knee bone or not.
- Osteoarthritis Detection:
  Validate the system's ability to differentiate a knee osteoarthritis from a normal knee bone accurately.
- Severity Classification:
  Test the severity classification functionality to ensure it correctly classifies the severity level of osteoarthritis.
- Treatment Recommendation:

Confirm that the system provides appropriate treatment recommendations based on the severity of detected osteoarthritis level.

Non – Functionality Testing Criteria

- Performance:
  Assess the system's speed in analyzing the image and providing the output. Also, system's responsiveness, particularly in relation to scalability.
- Reliability:
  Ensure that the system consistently delivers accurate results and performs reliably under varying conditions.
- Security:
  Ensure that user data, including uploaded images and extracted information, is processed securely to prevent unauthorized access or data leakage.
- Maintainability:
  Ensure that the code is well-structured and comprehensively documented to facilitate future updates, bug fixes, and enhancements.

8.4 Model Evaluation

The confusion matrix and classification report has been used to conduct model evaluations. The reason for selecting these two methods is because they provide detailed information about the effectiveness and performance of the models.

The confusion matrix shows a tabular overview of the model's predictions compared to the true labels for each class. The 4 key metrices are;

- True Positives (TP)
- True Negatives (TN)
- False Positives (FP)
- False Negatives (FN)

The classification report provides a comprehensive evaluation by calculating important performance measures, such as precision, recall, F1-Score and accuracy.

- Precision: the percentage of positive predictions that are correct.
- Recall: percentage of positive cases that are detected.
- F1-Score: harmonic mean of precision and recall.
- Accuracy: percentage of model correctly making predictions.

8.4.1 Detecting knee bone in x-ray

Convolutional Neural Network and ResNet-50 were used to train the model to determine whether the input image is a knee bone or not. An x-ray dataset with 5778 knee bones and 1708 non knee bones was used to train these models.

Table 16 Detecting Knee Bone- Model Testing

Model	**Description**	**Train Accuracy**	**Train Loss**	**Validation Accuracy**	**Validation Loss**	**Remark**
ResNet50	10 epoch	98.53%	0.049	97.30%	0.069	Overfitting
ResNet50	10 epoch(with Image Rotation)	97.25%	0.096	96.40%	0.103	Overfitting
ResNet50	20 epoch	98.33%	0.063	97.56%	0.066	Overfitting
CNN	5 epoch	97.42.%	0.085	98.52%	0.039	Could improve
CNN	10 epoch	98.92%	0.031	99.36%	0.018	

According to the table it can be seen that ResNet 50, a pre trained model, was not reliable because the validation loss is higher than the training loss, indicating potential overfitting. A CNN model was used and the model demonstrated consistent performance with good accuracy across both training and validation sets. Its architecture, which was customized to the features of the dataset, improved generalization and lessened the overfitting problems that the ResNet50 model possessed. As a result, switching to the CNN model produced better outcomes and increased dependability for our classification task.

Classification Report

```
Classification Report for Testing Dataset:
               precision    recall  f1-score   support

NonKnee bones       0.99      0.98      0.99       110
    Knee bone       0.99      0.99      0.99       141

     accuracy                           0.99       251
    macro avg       0.99      0.99      0.99       251
 weighted avg       0.99      0.99      0.99       251
```

Figure 19 Detecting Knee Bone- Classification Report

Confusion Matrix

The below results are from the test dataset which has 251 x-rays.

Table 17Detecting Knee Bone- Confusion Matrix

Metric	Description	Value
True Positive	Number of knee bones classified as knee bones.	140
True Negative	Number of non-knee bones classified as non-knee bones.	108
False Positive	Number of non-knee bones classified as knee bones.	2

False Negative	Number of knee bones classified as non-knee bones.	1

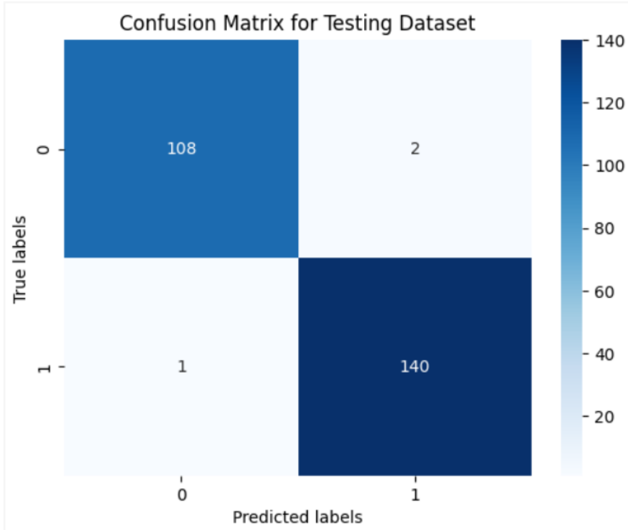

Figure 20 Detecting Knee Bone- Confusion Matrix

8.4.2 Detecting normal and diseased knee bones

The pre-trained models VGG-16 and MobileNetV2 were used to train the model to determine if a knee bone was normal or not. A dataset with 2268 normal bones and 3497 abnormal bones was used. However, the outcomes of these models showed that the model was underfitting since the test accuracy was higher than the train accuracy.

Table 18 Detecting Normal and Diseased Knee- Model Test

Model	Description	Train Accuracy	Test Accuracy	Remark
VGG-16(1)	No changes done to dataset, adaptive thresholding and contour extraction used for pre-processing.	55%	58%	Underfitting
VGG-16(2)	model was refined by using contrast enhancement, binary thresholding, edge detection using Canny and contour extraction when	50%	56%	Underfitting

	pre-processing the images.			
VGG-16(3)	Normal bone class was oversampled to balance the dataset.	40%	-	Low train accuracy.
MobileNetV2(1)	Same dataset and image pre-processing techniques.	50%	69%	Underfitting
MobileNetV2(2)	Dataset was refined and balanced.	50%	70%	Underfitting

According to the above table, it can be seen that none of the models were reliable to use in this project. Therefore, the dataset was altered and a dataset of 8370 was made. The dataset was not only divided in to train and test, it also included a validating set of images to further enhance the machine learning model. The following steps were taken to train the model:

- Images were rescaled by dividing pixel values by 255 to ensure they are in the range of [0,1].
- MobileNetV2 was used by excluding top classification layers to allow feature extraction.
- Global Average Pooling 2D layer is added to reduce spatial dimensions and flatten features. This helps reduce computational complexity.
- Two dense layers with 128 neurons each with ReLU activation and a dropout rate of 0.5 was added.
- Another dense layer with a single neuron and sigmoid activation function is used as the output layer for binary classification.

Finalized Model: Following the above alterations, the new model was trained, validated and tested and received an accuracy of 76%.

Classification Report

```
Classification Report:
              precision    recall  f1-score   support

      normal       0.74      0.81      0.77       639
    abnormal       0.79      0.72      0.75       639

    accuracy                           0.76      1278
   macro avg       0.76      0.76      0.76      1278
```

Figure 21Detecting Normal and Diseased Knee- Classification Report

Confusion Matrix

Table 19 Detecting Normal and Diseased Knee- Confusion Matrix

Metric	Value
True Positive	Number of normal bones classified as normal bones.
True Negative	Number of abnormal bones classified as abnormal bones.
False Positive	Number of normal bones classified as abnormal bones.
False Negative	Number of abnormal bones classified as normal bones.

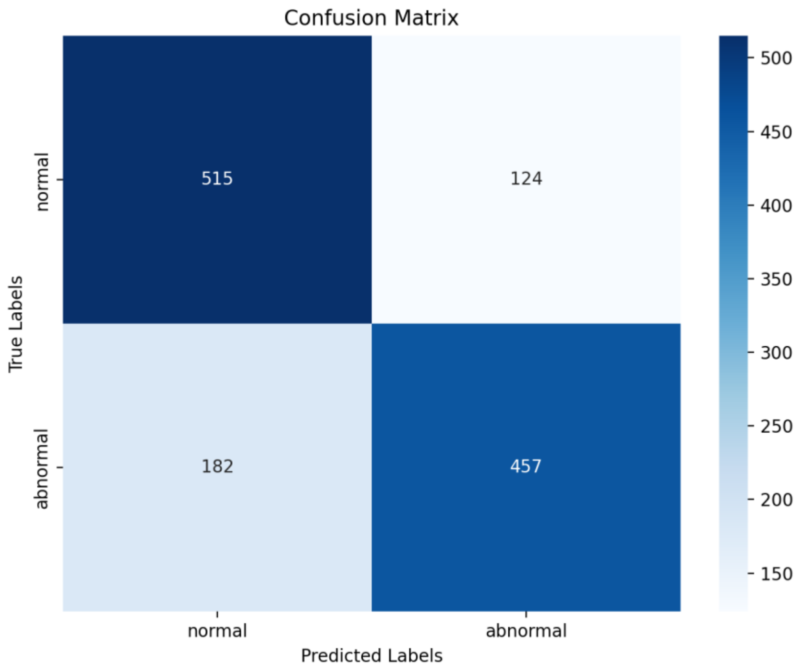

Figure 22 Detecting Normal and Diseased Knee- Confusion Matrix

8.4.3 Severity grading

A combination of custom Convolutional Neural Network, VGG16 and a custom neural network were used to train the model to classify the KL grade of the Osteoarthritis. A dataset of 5007 x-ray images of abnormal knee joints of 4 different grades was used.

Table 20 Severity Grading- Model Test

Model	**Description**	**Train Accuracy**	**Test Accuracy**	**Remark**
VGG16 + Custom Neural Network	10 epoch + 4 Kfold split	99.1%	51%	Overfitting
VGG16 + Custom	Kfold split inside each grade +	98.3%	47%	Overfitting

Neural Network	categorical encoder			
VGG16 + Custom CNN	100 epoch	87.7%	61%	Improved but overfitting

According to the table, overfitting was unavoidable therefore the highest possible test accuracy model was chosen. Even after combining all images from test, validate and train into one folder to get the maximum dataset size. During testing of these models it was also found that adding extra data preprocessing methods, data augmentation, and etc. seemed to reduce the model performance, therefore entailing a simple and minimalistic model to do the process the best. Therefore, in the finalized model a CNN with VGG16 pre-trained models were taken.

- The pre-trained VGG16 model was used to extract features from the images.
- All layers in the VGG16 model were frozen to be non-trainable.
- Resize the output of the base VGG16 model to match the input size expected by the custom CNN.
- Added a 1x1 Convolutional layer to reduce the number of channels to 3 to ensure the compatibility with the CNN input.
- Defined a custom CNN model using Sequential. It consists of several convolutional layers followed by max-pooling layers for feature extraction, and then fully connected layers for classification.

Finalized model: With the above alterations the latest model achieved 61% accuracy.

Classification Report

```
Classification Report:
              precision    recall  f1-score   support

           1       0.55      0.56      0.56       275
           2       0.60      0.68      0.63       403
           3       0.77      0.58      0.66       200
           4       0.65      0.55      0.59        44

    accuracy                           0.61       922
   macro avg       0.64      0.59      0.61       922
weighted avg       0.62      0.61      0.62       922
```

Figure 23 Severity Grading- Classification Report

Confusion Matrix

Table 21 Severity Grading- Confusion Matrix

Metric	Description	Value
True Positive	The number of severity level 1 bones classified as severity level 1	117
	The number of severity level 2 bones classified as severity level 2	210
	The number of severity level 3 bones classified as severity level 3	84
	The number of severity level 4 bones classified as severity level 4	26
True Negative	None	None
False Positive	The number of severity level 1 bones classified as severity level 2	20
	The number of severity level 2 bones classified as severity level 3	7
False Negatives	The number of severity level 2 bones classified as severity level 1	18
	The number of severity level 3 bones classified as severity level 2	12
	The number of severity level 4 bones classified as severity level 3	1
	The number of severity level 4 bones classified as severity level 1, 2, 3	1, 2, 1

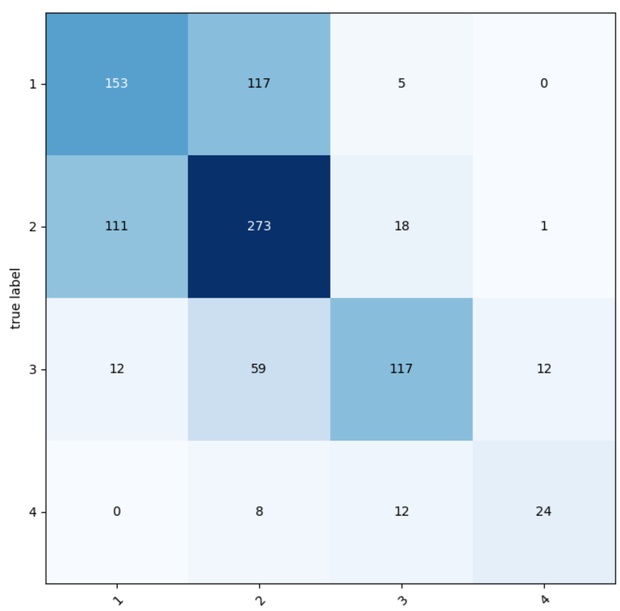

Figure 24 Severity Grading- Confusion Matrix

8.4.4 Treatment Recommendation

A random forest model was constructed to recommend treatments based on the severity grade of knee osteoarthritis. The dataset used for training consisted of 200 treatment instances. After training and evaluation, the model was deployed for practical use. System takes severity levels as input to receive tailored treatment recommendations. This test has been conducted under 20% of the data.

Table 22 Treatment Recommendation- Model Test

Model	Description	Train Accuracy	Test Accuracy	Remark
Random forest model	The dataset was not much organized as there were less columns in the dataset.	98%	60%	Overfitting
Random forest model	The dataset was organized by inserting several columns to it.	85%	90%	Underfitting

Finalized model: According to the above table it can be seen that the first model is overfitting as the train accuracy is higher than test accuracy and that model used a dataset which was not very organized. Therefore, another model was created and the dataset was modified to achieve an accuracy of 90% without the issue of underfitting or overfitting.

Classification report

```
              precision    recall  f1-score   support

           0       0.67      0.67      0.67         6
           1       0.71      0.71      0.71         7
           2       1.00      1.00      1.00         7
           3       1.00      1.00      1.00        14
           4       1.00      1.00      1.00         6

    accuracy                           0.90        40
   macro avg       0.88      0.88      0.88        40
weighted avg       0.90      0.90      0.90        40
```

Figure 25 Treatment Recommendation- Classification Report

Confusion matrix

Table 23 Treatment Recommendation- Confusion Matrix

Metric	Description	Value
True Positive	The number of severity level 0 bones classified as severity level 0.	4
True Negative	The number of severity level 1 bones classified as severity level 1.	5
	The number of severity level 2 bones classified as severity level 2.	7
	The number of severity level 3 bones classified as severity level 3.	14
	The number of severity level 4 bones classified as severity level 4.	6
False Positive	The number of severity level 0 bones classified as severity level 1.	2
	The number of severity level 1 bones classified as severity level 0.	2
False negative	None in this case	

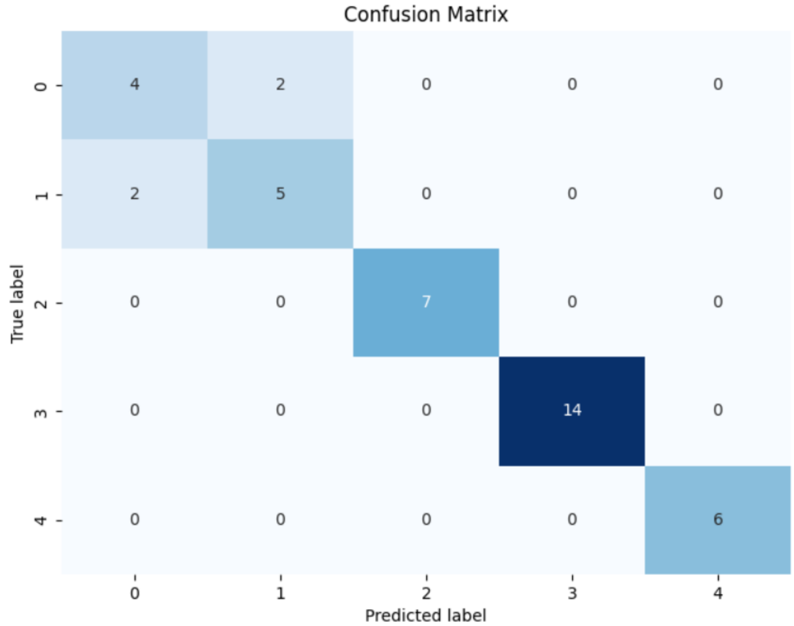

Figure 26 Treatment Recommendation- Confusion Matrix

8.5 Benchmarking

The purpose of benchmarking is to compare a product to similar products from leading companies to determine its effectiveness and identify areas for improvement. There are two kinds of benchmarking: competitive, which compares our product with those of top-level companies to see how well it performs, and technical, which evaluates our product's capabilities based on similar products from top-performing companies. As there are no proper products for this area made yet, research papers and articles were used to compare.

Table 24 Benchmarking

System Component	**Human Level & State-of-the-art-performance**	**Comparison of the model with similar products**
Detecting knee bone in x-ray	-	Most other products have not included a clear feature that checks whether the x-ray is of a knee joint, to do this CNN model with binary classification has been used to detect knee bones where other researchers have used KNNs and SVM.

Detecting normal and diseased knee bones	-	This model uses MobileNetV2 along with binary classification to determine if a knee bone is normal or not. If it is normal, the treatment recommendation model will recommend prevention suggestions and if diseased severity grading model will determine the severity of the disease. Other models have used Logistic Regression, DenseNet169, and DHL-II, however these model did not check if the knee bone was normal or not, instead it analyzed the abnormal knee.
Severity grading	Requires specialists and a lot of time to detect the severity by just reading an x-ray. This has a high risk of misdiagnosis. (Abdul Sami Mohammed et al., 2023)	The model for this component combines a pre-trained VGG16 base model with the CNN for feature extraction where other products may use different pre-trained models.
Treatment Recommendation	-	Treatment recommendation was a feature that was not found integrated to the detection system in other products. This component provides multiple suited non-medical treatments for the respective severity grading.

8.6 Functional Testing

Unit testing has been used to conduct functional testing.

Table 25 Functional Testing

Test Case	Description	Input	Expected Output	Actual Output	Remark

01	Read QR code, extract the image from it and preprocess for diagnosing.	QR Code	Image embedded in the QR Code	Image embedded in the QR Code	PASS
02	Extract features from X-ray to check if the x-ray is of a knee bone or not.	X-ray image	Determine if the x-ray is a knee bone or not.	“Not a knee bone”	PASS
				“Knee bone verified”	PASS
03	Osteoarthritis detection: detects of the knee bone is normal or not.	X-ray image of knee	Determine if the knee bone is normal or not.	“Normal”	PASS
				“Abnormal”	PASS
04	Severity assessment: detects the severity of the disease	Abnormal image of knee x-ray	Determine the severity of the abnormal knee bone	'Doubtful: KL grading- 1'	PASS
				'Minimal: KL grading- 2'	PASS
				'Moderate: KL grading- 3'	PASS
				'Extreme: KL grading- 4'	PASS
05	Treatment recommendation: treatment plans need to be recommended according to the severity of the disease	Severity of disease	Suitable treatment recommendations for each level.	Treatment plans are accurately displayed.	PASS

8.7 Module and Integration Testing

To integrate the knee osteoarthritis detection system, each component has been systematically tested, ensuring they function together seamlessly. By prioritizing critical modules, thoroughly testing interfaces, and using real-world data, the system's accuracy in detected the disease and recommending treatment options was validated. Additionally, security measures were implemented to ensure access controls, ensuring the system's reliability and compliance with privacy regulations. Given below are the steps of integration testing done for this project.

1. Test Independently: Check each part of the system independently first.
2. Start with important components: Focus on testing the most crucial parts early.
3. Check connections: Make sure different parts of the system can communicate with each other smoothly.
4. Use Real Data: Test with real-world examples to see if the system works as, it should.
5. Run tests and keep records: Conduct tests and write down results.
6. Scan for problems: Monitor the results to find any errors or issues.
7. Fix errors: Correct the issue which was identified.
8. Repeat Testing: Continue testing until everything works as planned.

8.8 Non- Functional Testing

8.8.1 Accuracy Testing

Accuracy refers to the measure of how often a model's predictions match the actual outcomes, expressed as the ratio of correctly predicted instances to the total instances. It indicates the system's overall correctness and effectiveness in classification or prediction tasks.
This system has a relatively good accuracy rate in detecting knee bones, normality of the knee, severity levels, and treatment recommendations, ensuring reliable and precise assessments for informed assistance in decision-making for doctors.

8.8.2 Performance Testing

Performance testing evaluates how well the system performs under different conditions, ensuring it meets response time and stability requirements.

Output Time: The knee osteoarthritis detection system aims to provide users with results within seconds of submitting the QR code for analysis, ensuring prompt access to diagnostic information.

8.8.3 Load Balancing

Through the equitable distribution of network traffic among several servers, load balancing maximizes resource efficiency and enhances system responsiveness. It distributes requests according to the health of the server and the strain on it, preventing server overload. Load balancing improves system speed and reliability by utilizing strategies like least connections and round-robin.

However, load balancing is not applicable for this system, since this system is hosted on a local server, where network traffic is not an affecting factor.

8.9 Limitations

During the testing process the below limitations were found -

- Processor usage for most of the machine learning models were high during training.
- The model for severity grading took more than 12 hours to train.
- Not saving the patient records after recommending treatments.
- Not providing personalized treatment plans with regard to allergies, age, gender and other factors.
- Non-medical treatments only.
- Accuracy of the system is around 75% which is not efficient enough for an important field such medicine, thus, this system can be used as an assistance in detecting knee osteoarthritis and recommending treatment. Doctors cannot fully rely on this system.

8.10 Chapter Summary

This chapter provides a thorough analysis of the experiments and tests carried out to confirm the project’s functionality and guarantee that it complies with the requirements and goals set. The system was tested using functionality testing, benchmarking, module and integration testing and non-functionality testing. Thorough tests were conducted to analyze the performance and accuracy of the entire system. Furthermore, limitations were discovered and addressed in this chapter.

References

Hsu, H. and Siwiec, R.M. (2023). Knee Osteoarthritis. [online] Nih.gov. Available at: https://www.ncbi.nlm.nih.gov/books/NBK507884/#:~:text=Knee%20osteoarthritis%20(OA)%2C%20also,two%20types%2C%20primary%20and%20secondary . [Accessed 24 Oct. 2023].

World (2023). Osteoarthritis. [online] Who.int. Available at: https://www.who.int/news-room/fact-sheets/detail/osteoarthritis#:~:text=Key%20facts,%25%20are%20female%20(1) . [Accessed 25 Oct. 2023].

Kotti, M., Duffell, L.D., A. Aldo Faisal and McGregor, A.H. (2017). Detecting knee osteoarthritis and its discriminating parameters using random forests. Medical Engineering & Physics, [online] 43, pp.19–29. doi:https://doi.org/10.1016/j.medengphy.2017.02.004 . [Accessed 25 Oct. 2023].

Abdullah, S.S. and Rajasekaran, M.P. (2022). Automatic detection and classification of knee osteoarthritis using deep learning approach. La radiologia medica. doi:https://doi.org/10.1007/s11547-022-01476-7. [Accessed 25 Oct. 2023].

TIULPIN, A. et al., 2018. Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach. Scientific Reports, 8(1), p. 1727. Available from: https://www.nature.com/articles/s41598-018-20132-7 . [Accessed 25 Oct. 2023].

Chen, P., Gao, L., Shi, X., Allen, K. and Yang, L. (2019). Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. Computerized Medical Imaging and Graphics, 75, pp.84–92. doi:https://doi.org/10.1016/j.compmedimag.2019.06.002 [Accessed 25 Oct. 2023].

KOKKOTIS, C. et al., 2020. Machine learning in knee osteoarthritis: A review. Osteoarthritis and Cartilage Open, 2(3), p. 100069. Available from: https://www.researchgate.net/publication/341144350_Machine_Learning_in_Knee_Osteoarthritis_A_Review . [Accessed 25 Oct. 2023].

NIAMS (2017). *NIAMS Health Information on Osteoarthritis*. [online] National Institute of Arthritis and Musculoskeletal and Skin Diseases. Available at: https://www.niams.nih.gov/health-topics/osteoarthritis/diagnosis-treatment-and-steps-to-take#:~:text=Early%20joint%20damage%20does%20not,is%20locking%20or%20giving%20out . [Accessed 25 Oct. 2023].

Tariq, T., Suhail, Z. and Nawaz, Z. (2023). Machine Learning Approaches for the Classification of Knee Osteoarthritis. [online] doi: https://ieeexplore.ieee.org/abstract/document/10252236 [Accessed 13 October 2023]

Neslihan Bayramoglu, Aleksei Tiulpin, Jukka Hirvasniemi, Nieminen, M.T. and Simo Saarakkala (2020). "Adaptive segmentation of knee radiographs for selecting the optimal ROI in texture analysis." ("Adaptive segmentation of knee radiographs for selecting the ... - PubMed") Osteoarthritis and Cartilage, [online] 28(7), pp.941–952. doi: https://www.sciencedirect.com/science/article/pii/S1063458420309481 [Accessed 13 October 2023]

Abdo, A.A., Wafa El-Tarhouni, Asma Fathi Abdulsalam and Abdelgader Bubaker Altajori (2022). Estimating the severity of knee osteoarthritis using Deep Convolutional Neural Network based on Contrast Limited Adaptive Histogram Equalization technique. [online] doi: https://ieeexplore.ieee.org/abstract/document/9914285 [Accessed 13 October 2023]

Sameh Abd El-Ghany, Elmogy, M. and A. A. Abd El-Aziz (2023). A fully automatic fine tuned deep learning model for knee osteoarthritis detection and progression analysis. Egyptian Informatics Journal, [online] 24(2), pp.229–240. doi: https://www.sciencedirect.com/science/article/pii/S1110866523000208 [Accessed 13 October 2023]

Sozan Mohammed Ahmed and Mstafa, R.J. (2022). Identifying Severity Grading of Knee Osteoarthritis from X-ray Images Using an Efficient Mixture of Deep Learning and Machine Learning Models. Diagnostics, [online] 12(12), pp.2939–2939. doi: https://www.mdpi.com/2075-4418/12/12/2939 [Accessed 13 October 2023]

Clinic, C. (2023). *Osteoarthritis: Symptoms, Causes & Treatment Options*. [online] Cleveland Clinic. Available at: https://my.clevelandclinic.org/health/diseases/5599-osteoarthritis [Accessed 8 Oct. 2023].

Doubleday, K., Zhou, J., Zhou, H. and Fu, H. (2021). Risk controlled decision trees and random forests for precision Medicine. Statistics in Medicine. doi: https://doi.org/10.1002/sim.9253 . [Accessed 15 Oct 2023]

ieeexplore.ieee.org. (n.d.). Ensemble Machine Learning (Grid Search & Random Forest) based Enhanced Medical Expert Recommendation System for Diabetes Mellitus Prediction. [online]

Available at: https://ieeexplore.ieee.org/abstract/document/9885312/figures#figures [Accessed 14 Oct. 2023]

Doubleday, K., Zhou, H., Fu, H. and Zhou, J. (2018). An Algorithm for Generating Individualized Treatment Decision Trees and Random Forests. Journal of Computational and Graphical Statistics, 27(4), pp.849–860. doi: https://doi.org/10.1080/10618600.2018.1451337

ieeexplore.ieee.org. (n.d.). Predictions of Diabetes and Diet Recommendation System for Diabetic Patients using Machine Learning Techniques. [online] Available at: https://ieeexplore.ieee.org/abstract/document/9456365 [Accessed 14 Oct. 2023].

Antony, J., McGuinness, K., Moran, K. and O'Connor, N.E. (2017). Automatic Detection of Knee Joints and Quantification of Knee Osteoarthritis Severity Using Convolutional Neural Networks. Lecture Notes in Computer Science, [online] pp.376–390. doi: https://doi.org/10.1007/978-3-319-62416-7_27 [Accessed 14 Oct. 2023].

Zebari, D. A., Sadiq, S. S., & Sulaiman, D. M. (2022). Knee Osteoarthritis Detection Using Deep Feature Based on Convolutional Neural Network. In 2022 International Conference on Computer Science and Software Engineering (CSASE), Duhok, Iraq (pp. 259-264). doi: 10.1109/CSASE51777.2022.9759799. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9759799&isnumber=975945 [Accessed 14 Oct. 2023]

Antony, J., McGuinness, K., O'Connor, N. E., & Moran, K. (2016). Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks. In 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico (pp. 1195-1200). doi: 10.1109/ICPR.2016.7899799. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7899799&isnumber=789956. [Accessed 14 Oct. 2023]

Schwartz, A. J., Clarke, H. D., Spangehl, M. J., Bingham, J. S., Etzioni, D. A., & Neville, M. R. (2020). Can a Convolutional Neural Network Classify Knee Osteoarthritis on Plain

Radiographs as Accurately as Fellowship-Trained Knee Arthroplasty Surgeons? Retrieved from https://www.sciencedirect.com/science/article/pii/S0883540320304472 [Accessed 15 October 2023]

DONG, Y. et al., 2023. Classification of knee osteoarthritis based on quantum-to-classical transfer learning. *Frontiers in Physics*, 11, p. 1212373. Available from: https://www.frontiersin.org/articles/10.3389/fphy.2023.1212373/full

2024. [online]. Techscience.com. Available from: https://www.techscience.com/cmc/v70n3/44972/html [Accessed 6 February 2024].

2024. [online]. Ieee.org. Available from: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10179858 [Accessed 6 February 2024].

IJRASET, 2024. *Deep Neural Network-based Knee Osteoarthritis Grading Using X-Rays*. [online]. Ijraset.com. Available from: https://www.ijraset.com/research-paper/deep-neural-network-based-knee-osteoarthritis-grading-using-x-rays [Accessed 6 February 2024].

Abdul Sami Mohammed, Ahmed Abul Hasanaath, Latif, G. and Bashar, A. (2023). Knee Osteoarthritis Detection and Severity Classification Using Residual Neural Networks on Preprocessed X-ray Images. Diagnostics, [online] 13(8), pp.1380–1380. doi: https://doi.org/10.3390/diagnostics13081380 . [Accessed 17 March 2024]