Universitat Oberta de Catalunya (UOC)

Master's Degree in Data Science

FINAL MASTER THESIS

Area of Medicine

Automatic Detection of Knee Joints and Classification of Knee Osteoarthritis Severity from Plain Radiographs using CNNs

Author: David Durán Olivar

Supervisor: Héctor Espinós Morató

Supervisor: Darwin Patricio Castillo Malla

Professor: Jordi Casas Roma

Karlsruhe (Germany), January 2022

Copyright

[1] https://github.com/d-duran/KOA-location-diagnose-CNN

MASTER THESIS PROFILE

Thesis title:	Automatic Detection of Knee Joints and Classification of Knee Osteoarthritis Severity from Plain Radiographs using CNNs
Name of the author:	David Durán Olivar
Name of the Associate Professor:	Ph.D Héctor Espinós Morato
PRA:	Ph.D Jordi Casas Roma
Delivery date:	01/2022
Master's Degree:	Data Science
Master Thesis area:	Area of Medicine
Language:	English
Keywords	Knee Osteoarthritis, Convolutional Neural Networks, U-Net, X-Rays

A Silvia, por su apoyo incansable.

Abstract

Knee Osteoarthritis (OA) is the most common type of arthritis and it is typically the result of wear and tear, and progressive loss of articular cartilage which may eventually lead to disability. OA diagnosis is typically conducted by performing a physical examination of the knee by means of a visual inspection of radiographic imaging. Based on the presence of OA pathological features of joint space narrowing, osteophyte formation or sclerosis, Kellgren-Lawrence (KL) system is typically used to classify the severity of the disease into one of five ranked grades. The conclusion regarding the presence and severity of knee OA may differ due to the subjective nature of the assessment.

In this study, we present a computer-aided diagnosis method based on Convolutional Neural Networks (CNN) to automatically locate and score knee OA severity from X-ray images according to the KL grading scale. Location of the knee joints is achieved by considering a region of interest (ROI) segmentation with U-Net architecture. Transfer learning from pre-trained CNN architectures is considered for knee OA severity assessment.

Our method yields a quadratic Cohen Kappa coefficient of 0.87 and a weighted average f1-score of 72%. In addition, we show attention maps highlighting the strongest contribution to the network prediction. The visualization provides practitioners with information to assist in the diagnosis decision-making processes.

We conclude that our methodology achieves high accuracy in localizing knee joints out of plain X-ray images, as well as a very good performance in the OA diagnostic assessment of KL grades.

Keywords: Knee Osteoarthritis, Convolutional Neural Networks, U-Net, X-Rays

Resumen

La osteoartritis (OA) de rodilla es el tipo de artritis más común y es típicamente el resultado del deterioro y pérdida progresiva del cartílago articular, que puede conllevar discapacidad. El diagnóstico de la OA se realiza mediante una exploración visual de una radiografía de rodilla. En base a la presencia de las patologías asociadas a la OA como son el estrechamiento de la articulación, la formación de osteofitos o la esclerosis, se utiliza el sistema Kellgren-Lawrence (KL) para clasificar la severidad de la enfermedad en una escala de cinco grados. La conclusión sobre la existencia y severidad de la OA de rodilla puede diferir debido a la naturaleza subjectiva de la valoración.

En este estudio, presentamos métodos de diagnósticos asistidos basados en redes neuronales convolucionales (CNN) para localizar y diagnosticar automáticamente la severidad de la OA de rodilla presente en una radiografía de acuerdo al sistema KL. La localización de la articulación se consigue mediante la segmentación de una región de interés (ROI) con una arquitectura U-Net. Para la evaluación de la severidad de la OA de rodilla, se aplica la técnica de *transfer learning* sobre una serie de arquitecturas CNN pre-entrenadas.

Nuestro método obtiene un coeficiente Cohen Kappa cuadrático de 0.87 y una media ponderada de f1-score del 72%. Adicionalmente, se muestran mapas de atención que resaltan las regiones de la imagen con mayor contribución a la predicción de la red. La visualización ofrece información para ayudar en el proceso de evaluación del diagnóstico.

Se concluye que nuestro modelo alcanza una alta precisión en la localización de la articulación sobre radiografías, además de un rendimiento muy bueno en el diagnóstico de OA de rodilla según el sistema KL.

Keywords: Knee Osteoarthritis, Convolutional Neural Networks, U-Net, X-Rays

Contents

List of Figures

List of Tables

Acronyms

CAM Class Activation Map.

CE Crossentropy.

CLAHE Contrast Limited Adaptive Histogram Equalization.

CNN Convolutional Neural Network.

FCN Fully Convolutional Network.

GAP Global Average Pooling.

HOG Histogram of Oriented Gradients.

JI Jaccard index.

JSN Joint Space Narrowing.

KL Kellgren-Lawrence.

KPI Key Performance Indicator.

MRI Magnetic Resonance Imaging.

NDA NIMH Data Archive.

OA Osteoarthritis.

OAI Osteoarthritis Initiative.

PSD Power Spectral Density.

ROI Region of Interest.

SVM Support Vector Machine.

Chapter 1

Introduction

1.1 Project description and justification

Knee Osteoarthritis (OA), also known as degenerative joint disease, is the most common type of arthritis diagnosed and it is typically the result of wear and tear, and progressive loss of articular cartilage, being most common in elderly people [1]. The intensity of the clinical symptoms may vary for each individual; however, they typically become more severe, more frequent, and more debilitating over time, and may eventually lead to disability.

Patients typically present knee pain; therefore, it is essential to obtain a detailed history of their symptoms and to perform a physical examination of the knee by means of a visual inspection. Radiographic imaging is considered the standard gold in clinical and epidemiological analyses [2] and it is recommended as the major pathological features for OA could be present in the image, i.e. joint space narrowing (JSN), osteophyte formation, and sclerosis.

Based on these indicators, **Kellgren-Lawrence** (KL) [3] system is typically used in order to assess the severity of the disease. In this system, individual joints are classified into one of five grades, with 0 representing normal and 4 being the most severe disease. Each KL grade presents typical radiographic OA symptoms that are summarized in Table 1.1.

The conclusion regarding the presence and severity of knee OA may differ due to the subjective nature of the assessment. Medical experts require a considerable amount of knowledge and experience to make a valid diagnosis. Hence, computer-based tools for knee OA detection based on X-ray images, while not meant to replace a human expert, can serve as a **decision-**

supporting tool.

Table 1.1: Radiographic Osteoarthritis symptoms according to KL grade system

KL grade	Radiographic Osteoarthritis Symptoms
0	Healthy - None
1	Doubtful JSN and possible osteophytic lipping
2	Definite osteophytes and possible JSN on anteroposterior weight-bearing radiograph
3	Multiple osteophytes, definite JSN, sclerosis, and possible bony deformity
4	Large osteophytes, marked JSN, severe sclerosis, and definite bony deformity

In this work, I will propose a model for automatic detection and quantification of knee OA severity using CNNs. I will approach this work as an **image classification problem**, assessing knee OA severity out of plain X-ray images. This process involves two main steps:

1. Automatically detecting and extracting the region of interest (ROI), i.e. localizing the knee joints out of X-ray images.

2. Classifying the localized knee OA severity by assigning a KL grade.

Deep learning models automatically learn the relevant features to produce an output; however, the learnt features may not be relevant to the disease as the models may react to background noise or image artifacts [4]. It is important to examine the region where the network is detecting the disease-relevant features. Therefore, should the time constrain allow it, I will additionally define a methodology to obtain the class-discriminating activation maps which will allow to examine where the model is focusing its attention in the classification process.

In order to use the model to obtain relevant insights, it needs to be deployed in a manner that allows non-technical users to gain value from the previous work. Time permitting, I will additionally deploy the model into a functional application which would take a knee X-ray image as input, and would output a KL grade to support the expert in the diagnosis decision-making process.

1.2 Personal motivation

One of the big decisions we need to face in life is which path to follow after school, no matter whether the choice is the right one. Sometimes, one is not mature enough nor has the information required to make a conscious decision. Sometimes, we choose a path hoping to find a bright future, but eventually just find dissatisfaction.

After getting my degree in Mechanical Engineering, I mainly got my working experience in the aeronautic industry, specifically in aircraft structure analysis. However, what I thought would be the perfect job for me, turned out to be something that I did not completely enjoy.

Luckily, I heard of optimization algorithms, some of them based on Neural Networks, in a course I took at college, and that became a topic of interest for me. Some years after that, I decided that I needed a change, so I got the courage to step a leap forward onto a path I hope will lead to a new career full of opportunities.

Times are changing, and it is important to be part of that change in the most honest way possible. This project is framed in the area of Medicine, which I consider to be one of the better fields I could contribute to. By using Deep Learning, I intend to develop tools that may help in the early diagnosis of knee osteoarthritis, so that patients quality of life is as best as it can be.

1.3 Objectives definition

The **main objective** of this work is to implement a CNN-based methodology for Knee Osteoarthritis severity assessment based on the disease-relevant features present in knee radiographic images.

The **secondary objectives** of this work define the understanding of the domain of knowledge studied and the techniques used. Each of these objectives will be considered met when a key performance indicator (KPI) is met.

- Design and implementation of a CNN architecture to automatically detect and extract the ROI of a knee X-ray image.
  KPI: Functional scripts capable of generating cropped images of the ROI.

- Design and implementation of a CNN architecture to assign a OA severity grade to a knee X-ray image.
  KPI: Functional scripts capable of generating a prediction label of the localized ROI.

- Evaluation and interpretation of the classification output against the ground truth.

KPI: Selection and justification of the best metric, and assessment of the model performance.

- Define a methodology to obtain the class-discriminating activation maps in order to examine the region where the network is detecting the disease-relevant features.
  KPI: Being able to plot the activation maps obtained.

The code generated for the different implementations can be found in a dedicated GitHub repository created for this study[1].

1.4 Methodology

The main work of this project involves applying some type of analytical process to data (i.e. X-ray images) to derive insight from it. The initial dataset may be subject to change in terms of size, quality or availability; hence, a **dynamic data environment** scenario must be considered.

Only after the first analyses and data exploration are completed, will it be possible to define the scope of the problem and the line of action to accomplish the objectives. This scenario requires a flexible and agile working process in order to accommodate the **changing environment**.

This project is subject to a **time constraint**, therefore we will face the challenge of developing and releasing the work in a staged manner so that the progress is efficiently tracked and reported.

Since this work is defined by its **academic research nature**, its ultimate goal is to contribute with reproducible results and easily explained knowledge for the community to be able to validate it.

For all these reasons, **Guerrilla Analytics** [5] is considered to be the most appropriate data mining methodology for this project.

The basic Guerrilla Analytics workflow, as shown in Figure 1.1, is based on the following data mining flow: Domain and Data understanding, Data preparation, Modeling, Evaluation, and Deployment. There is no major differences from other data mining methodologies except

[1] https://github.com/d-duran/KOA-location-diagnose-CNN

for the fact that any disruption is accommodated in the working process in a manner that the project becomes iterative.

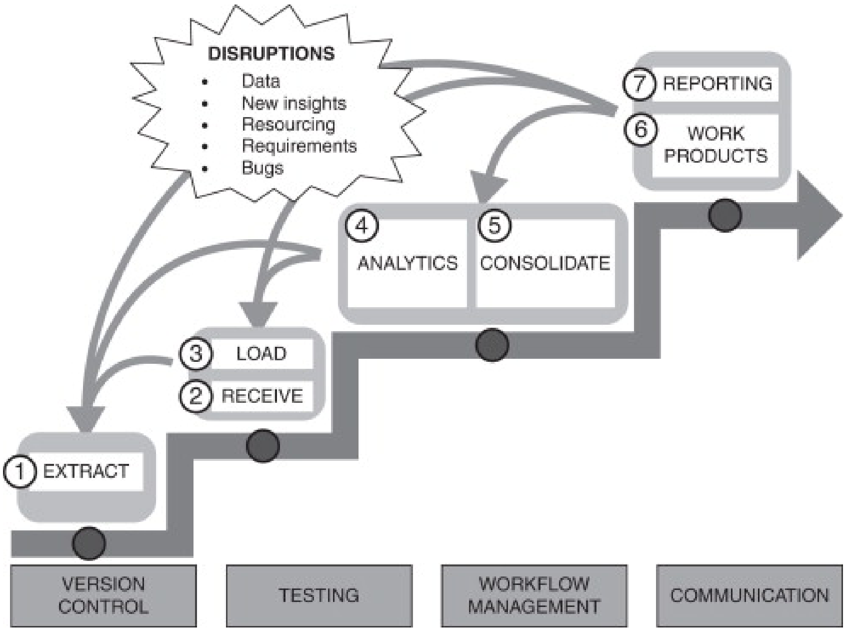


Figure 1.1: The Guerrilla Analytics Workflow (extracted from [5])

1.5 Planning

This section presents the different task that will be performed throughout the project using a **Gantt chart**. Figure 1.2 shows the detailed schedule where the main tasks correspond to the partial deliveries as per the university planning. These main tasks define the project milestones, including the final presentation and defense.

All main tasks are breakdown in order to define the work-specific tasks, which may be subject to minor modifications throughout the project.

1. **Project definition and scope**

   This chapter serves as an introduction to the Master Thesis topic, where the scope and objectives are presented, along with the personal motivation of the author and the methodology that will be followed.

2. **State of the art**

A through research will be carried out on published literature and studies in the field of knee osteoarthritis. This research will cover the understanding of the disease, its context and what are the most common detection methodologies. Finally. I will research on how the detection methodologies intertwine with the field of Deep Learning by focusing on how other researchers faced the detection and classification of knee OA. This will be the seed for defining my own research line.

3. **Project development**

   This is the main part of the thesis as it covers the development of the project. This step is split into different sub-tasks:

   - Analysis of the available data, which consist of several X-ray images.
   - Development of the Python scripts to create the required models, i.e. a model to automatically detect and extract ROIs, and a model for classification of the localized knee assigning a KL grade.
   - Training of the developed models.
   - Evaluation of the results after the training step.
   - Development of a class-discriminating activation map that will help on the model performance assessment.

4. **Report preparation**

   This step covers the documentation of the project development and the review process.

5. **Project presentation and public defense**

   As final step, the project will be presented in public for a tribunal to evaluate.

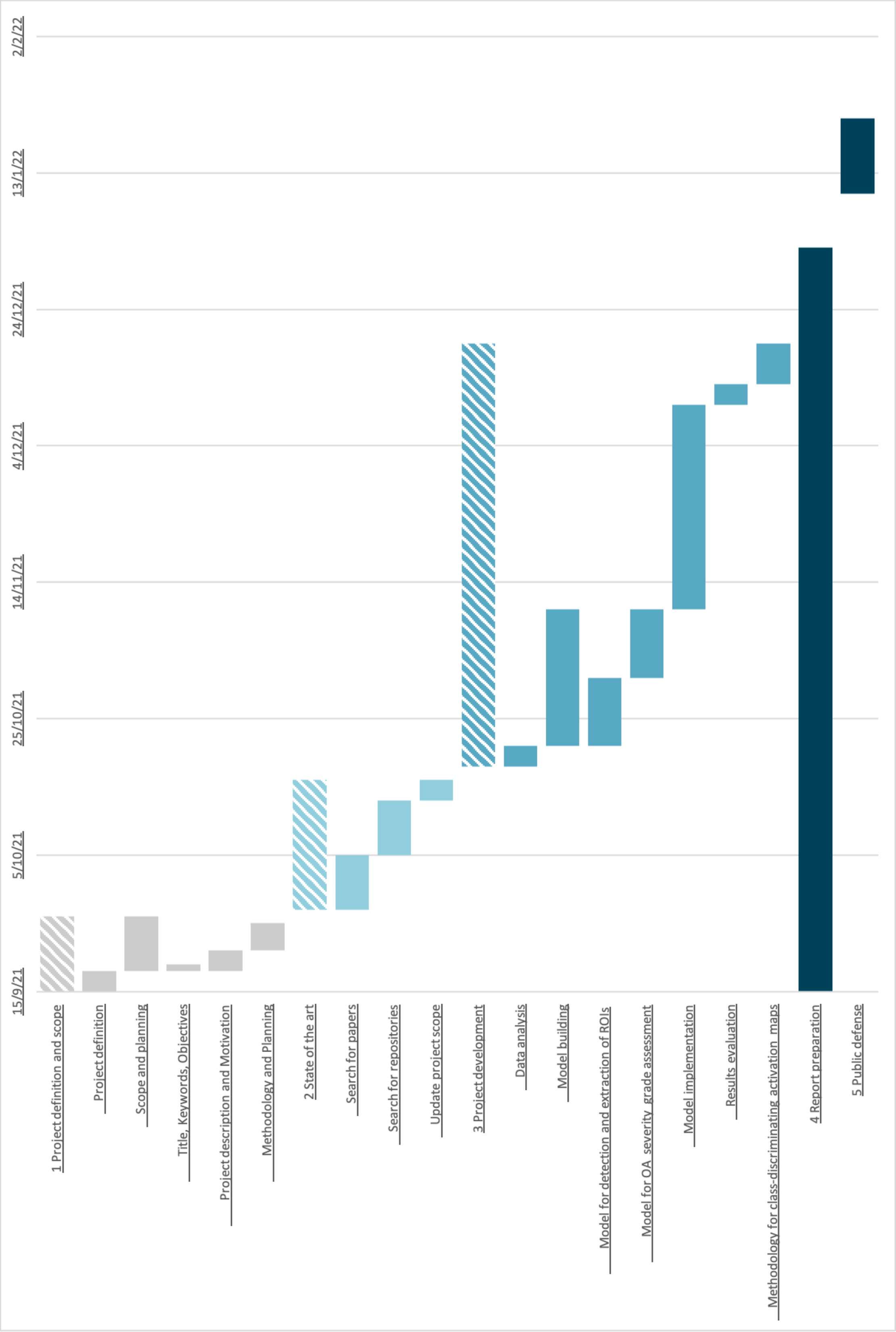


Figure 1.2: Gantt chart of the project schedule

Chapter 2

State of the art

2.1 Knee Osteoarthritis: definition, prevalence, incidence and impact

Knee Osteoarthritis (OA) is a long-term chronic disease characterized by the damage and loss of articular cartilage and many of its surrounding tissues. Although still under investigation, it is accepted that the origin of OA is multi-factorial, i.e. both inflammatory and biomechanical processes play an important role in the disease [6].

While knee OA is related to ageing, it is also associated with a variety of both modifiable and non-modifiable risk factors, including: obesity, family history, diabetes, systemic inflammatory mediators, joint shape and dysplasia, trauma, bone density, occupational injury, and gender [6].

There have been research conducted into the potential link between occupation and the risk of developing knee OA. Individuals whose knees have been subjected to significant loading of the joint are far more likely to develop OA [7], with obesity adding further risk.

A., Cui *et. al.* [8] quantitatively synthesized the published epidemiological data of knee OA in order to estimate the global/regional prevalence and incidence of knee OA. The authors included any publication that provided meaningful data from 1990-2000's to 2020; thus, capturing the changes of demographic structure and lifestyle, as epidemiological data have also changed. The study shows that the global **prevalence** of knee OA was 16.0% in individuals over 15 years old, and 22.9% over 40 years old. As regards of **incidence**, there was 203 per

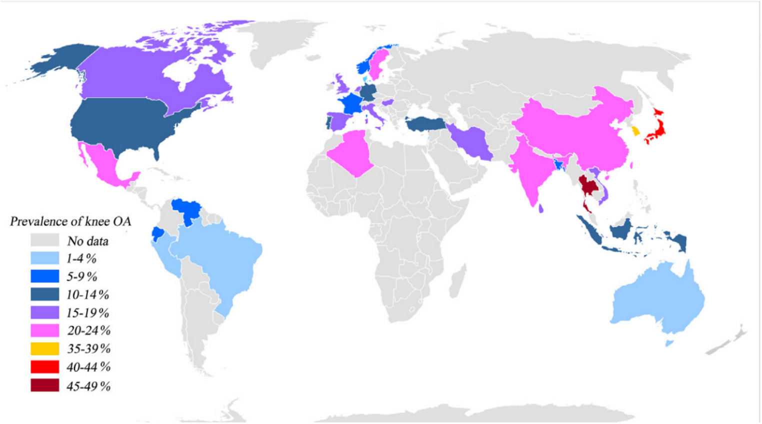


Figure 2.1: Global prevalence of knee OA per country (extracted from [8])

10,000 person-years in individuals over 20 years old. In 2020, there are an estimate of 86.7 million individuals (20 years or older) with incident knee OA worldwide. Figure 2.1 shows the global prevalence of knee OA per individual country, ranging from 1.6% to 46.3%. Moreover, as part of the Global Burden of Disease 2010 [9], it was found that, globally, of the 291 conditions considered, knee OA was ranked as the 11th highest contributor to global disability.

J. Salmon *et. al.* [10] carried out a systematic publication review in order to focus on the cost of hip and knee OA. The authors included published literature that covered the burden of lower-limb OA worldwide in terms of the economical impact. The study provides the mean annual cost per patient (converted to 2013 Euro considering Consumer Price Index of the countries), resulting in a **total cost** estimation of 0.7-12.0 k€/year per patient, with an average annual **direct cost** of 6.8 k€/year per patient. Direct costs vary according to surgery consideration, with average annual direct costs per patient without surgery, with surgery and awaiting surgery of 6.7 k€/year, 3 k€/year and 7.4 k€/year, respectively.

Although there is a substantial heterogeneity in how the costs have been estimated during the last decades, the total costs induced by knee OA could be as high as 408-817 billion €/year in Europe [10]. The burden of knee OA to society is high and, with life expectancy and obesity of the world's population gradually increasing, the implications of knee OA are not to be underestimated.

2.2 Knee Osteoarthritis clinical imaging

Knee OA is characterized by cartilage degradation and bone change, becoming more severe, more frequent, and more debilitating over time. Symptoms such as knee pain, stiffness and swelling are typically present in patients suffering from OA. Since it may eventually lead to disability, the detection of OA is relevant since early treatment could prevent cartilage and bone loss.

Medical imaging examination is usually performed in order to confirm OA diagnosis, to determine the involved parts of the joint, and to evaluate the stage of the disease. The ideal imaging modality for the assessment of OA should provide data pertaining to all joint structures, i.e. it should include a direct measure of both cartilage and bone.

Magnetic Resonance Imaging (MRI) is a technique in radiology that uses strong magnetic fields, magnetic field gradients and radio waves to generate images of the organs. Since this technique provides images with a high contrast of soft-tissues, they enable the direct assessment of cartilage as the main measure in the joint space narrowing phenomenon. Different MRI output measures of cartilage are also considered in the assessment of OA such as morphological properties (thickness, volume or surface defects) and biochemical composition [11].

MRI is a non-invasive and reliable technique that provides a whole view of the knee joint; however, it is indeed an expensive and lengthy medical procedure. Although MRI imaging modality is widely used in developed nations, the utilization of MRI is limited in most developing countries due to acquisition costs, lack of infrastructure and the expertise required for maintaining and running the systems.

Ultrasonography is a procedure that uses high-energy sound waves to look at tissues and organs inside the body. As a result, images of the tissues can be obtained from the echoes of sound waves. With this method, articular cartilaginous and periarticular soft-tissues structures can partially be visualized, and it is also a good method to assess inflammatory changes in knee OA. This may be helpful for follow-up of the disease during therapy as well as to detect early knee OA [2]. However, the main weakness of ultrasonography is the impossibility to look at deeper structures and its limited reproducibility, making this technique a supporting methodology to MRI or radiography.

Radiographic imaging (X-ray) is a modality that produces 2-dimensional images using radiation which is projected towards the body part of interest. When exposed to radiation, denser anatomy has a higher rate of radiation absorption than anatomy that is less dense. Therefore, the image receptor will receive more remnant radiation from soft-tissue areas and, conversely, denser parts such as bone will underexpose the image.

In the context of knee OA detection, when examining knee X-ray images, the practitioner should search for any of the pathological changes seen in OA joints, i.e. degradation of articular cartilage resulting in joint space narrowing, formation of osteophytes and changes in the subchondral bone. Osteophytes refer to bony lumps that grow around the joint (Figure 2.2), whereas subchondral bone refers to the bone lying immediately beneath the cartilage, and can be separated into two distinct anatomic entities: subchondral bone plate and subchondral trabecular bone (Figure 2.4). Based on the indicators listed above, Kellgren-Lawrence system (KL) [3] allows the OA severity assessment, classifying individual joints into one of five grades, with 0 representing normal and 4 being the most severe disease. Figure 2.2 shows four anterior-posterior X-ray samples of a left knee, each one corresponding to a different KL grade. In Figure 2.2-(a), doubtful joint space narrowing (JSN) and possible osteophytic lipping is indicated by the arrow, while (b) shows definite osteophytes and posible JSN indicated by the arrow. In Figure 2.2 (c), the arrows show multiple osteophytes, definite JSN, sclerosis and possible bony deformity; and, in (d), the image shows an osteophyte (right arrow), marked JSN (left arrow) and severe bone sclerosis (asterisk).

As shown in Figure 2.2, X-ray is limited by its 2-dimensional nature in the bone structure representation, its size and the cartilage defects as it cannot identify 3-dimensional changes in the articular structures. Despite these limitations, X-ray provides an indirect measure of articular cartilage by means of an assessment of the radiological joint space. Moreover, depending on the severity of the disease, should osteophytes or sclerosis be present, they would be visible to the expert in the X-ray image.

X-ray imaging is, perhaps, the most versatile and accessible of the imaging techniques in terms of procedure time involved and availability. It has been traditionally used as a cheaper alternative to assess OA disease compared to other techniques, despite the limitations associated with this approach. Radiography is considered the standard gold methodology when related to knee OA diagnosis and, therefore, knee OA detection using X-ray will be explored in detail in this study.

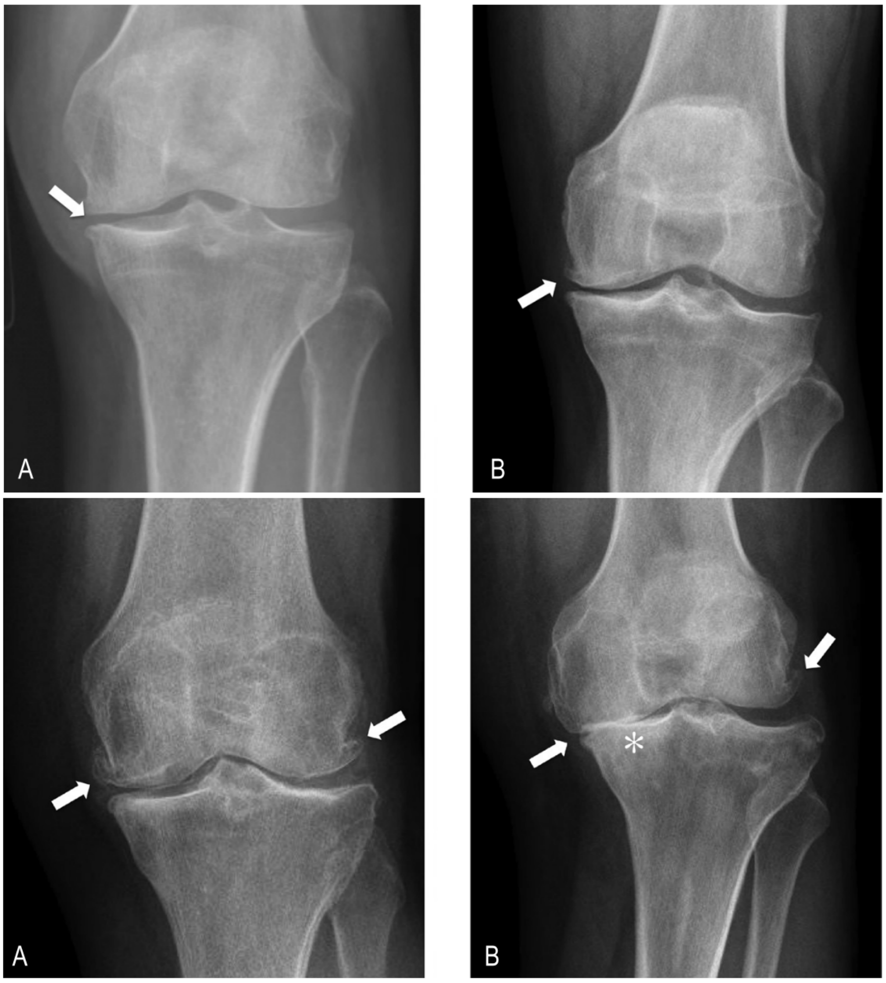


Figure 2.2: Anterior-posterior radiograph of a left knee with (a) mild OA (KL grade 1), (b) moderate OA (KL grade 2), (c) moderate to severe OA (KL grade 3) and (d) severe OA (KL grade 4). Extracted from [6].

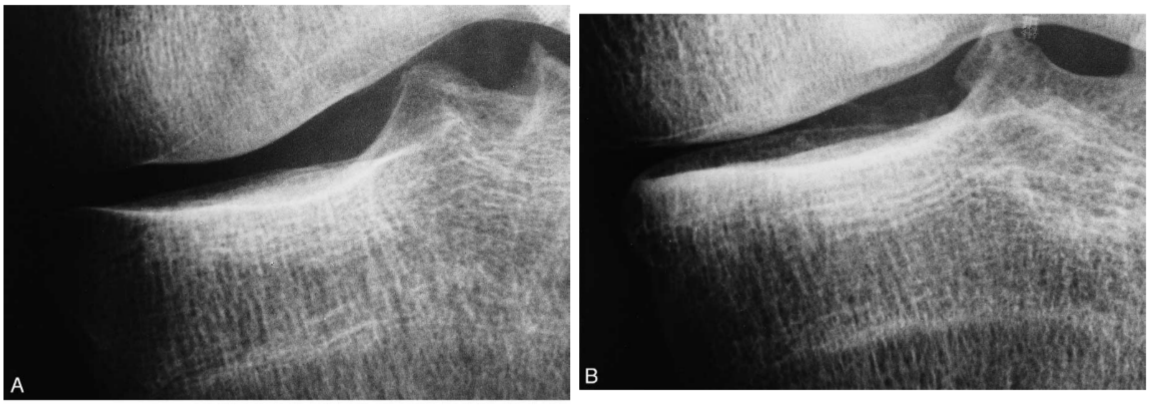


Figure 2.3: Medial compartment of OA knee with (a) early disease and (b) definite disease. Knee (b) shows the increase of thickness of the subchondral cortical plate and subjacent horizontal trabeculae resulting in a ladder-like appearance (extracted from [12]).

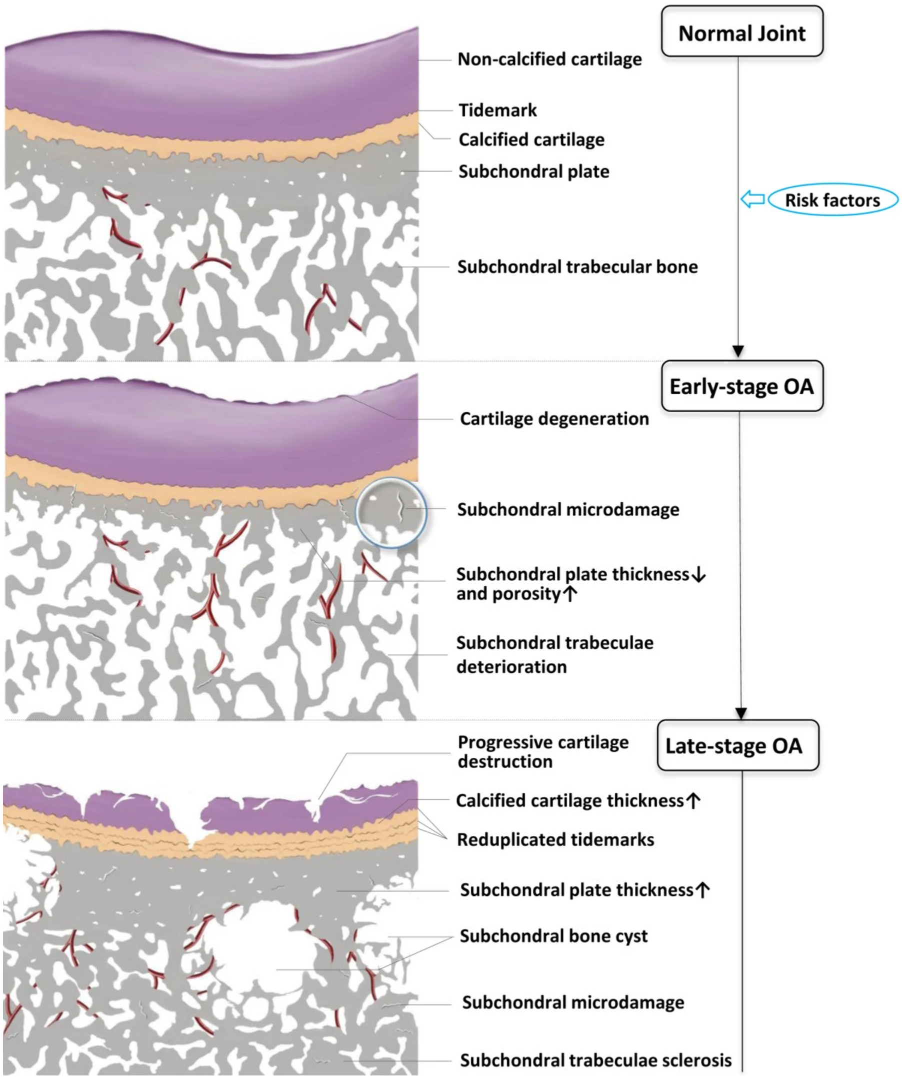


Figure 2.4: Model representation of Knee Osteoarthritis at different stages (extracted from [13])

2.3 Knee Osteoarthritis detection using X-ray

Evidences suggest that changes in bone occur early in the development of OA prior to the detection of radiographic abnormalities of JSN and osteophytes formation [12]. The **earliest changes** detected are elevated bone remodeling and subchondral bone loss, considered as a determinant of OA progression [13]. In the **late stage** of OA, increase in the thickness of the subchondral bone and subjacent horizontal trabeculae is detected, i.e. changes in trabecular structure: decrease of trabecular separation and transformation of trabeculae from rod-like into plate-like. These changes lead to a less stiff and dense bone, which is mechanically weaker.

2.3.1 Computer-aided diagnosis methods

Different computer-aided approaches to detection and analysis of OA using X-ray knee images have been proposed in the literature. Since significant changes occur to knee bone at different stages of the disease, some of these studies focus on the **analysis of the knee bone texture** as valuable information may be extracted from the bone structural patterns.

G.W. Stachowiak *et. al.* [14] presented a method to analyze the trabecular bone texture based on **fractal analysis** of the region of interest (ROI). The method calculates a set of fractal dimensions of the ROI in order to study the characteristics of the bone surface. This allows to quantify the roughness, the degree of surface anisotropy and direction of anisotropy. Then, a logistic regression model based on the parameters evaluates the prediction of OA progression.

In [15], A. Brahim *et. al.*. proposed a **spectral analysis** approach as the aided diagnosis method where the images are considered in the frequency domain instead of the spatial domain. Power Spectral Density (PSD) over the ROI lines was used in order to display the power of the variations in the image as a function of frequency. This analysis provides the mean periodogram that presents two different regimes with different slopes, separated by a cut-off frequency (see Figure 2.5-bottom). The cut-off frequency was then utilized as a high-pass filtering threshold in order to enhance the high frequencies of the image (associated with changing signals, i.e. texture border) before returning to the spatial domain (see Figure 2.5-top). Then, Naive Bayes and random forest classifiers were used for the classification task.

What stands out from these methods is that they perform a very detailed analysis of particular ROI of the knee articulation, focusing on features and patterns that may not be visible to the human eye. These methods, however, may be considerably affected by image noise, blur or resolution; therefore, image processing may be required in order to get reliable results.

2.3.2 Deep Learning Based Diagnosis Methods

Besides the changes in the bone that occur during OA, other radiographic features may be observed in X-ray images at different stages. Joint space narrowing and osteophytes formation are typically identified by the practitioner in order to diagnose and assess the severity of the disease. The assessment is traditionally approached as a classification problem by detecting these radiographic features, being the KL grades the ground thruth. In this line, some studies focus on developing methodologies that replicate the process that an expert would follow when

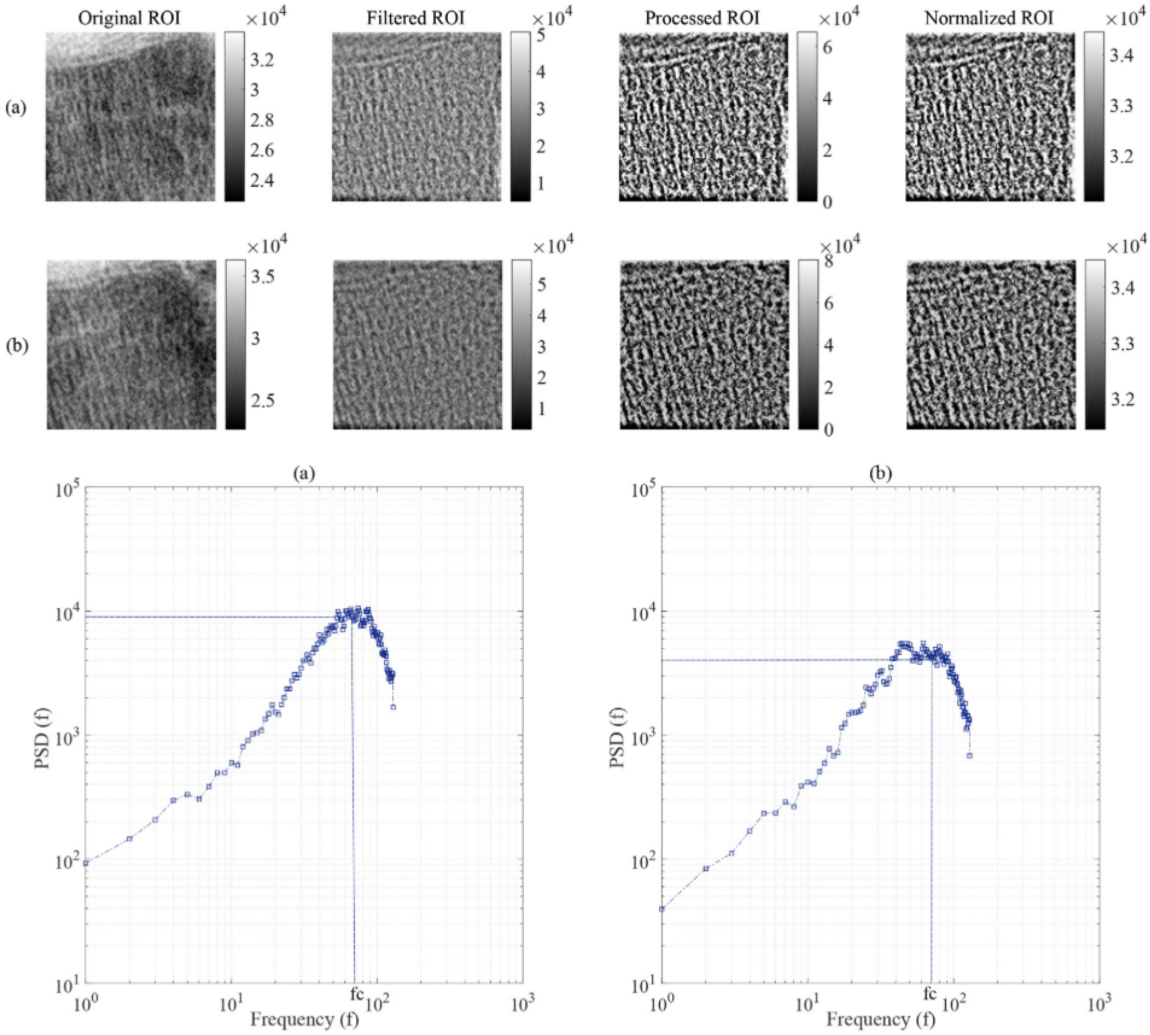


Figure 2.5: A representative ROI of the tibial trabecular bone (top) and the spectral transformations from (a) healthy subject and (b) a OA patient. Two representative periodograms (bottom) computed over the ROI lines for a (a) healthy subject and (b) a OA patient (extracted from [16])

examining knee X-ray images, i.e. extracting the relevant features of the images and use them to make a decision.

Recently, **convolutional neural networks** (CNNs) have outperformed other methods based on feature analysis, and they are highly successful in many computer vision tasks such as image recognition, automatic detection, segmentation, etc. Feature learning approaches such as CNNs exploits the 2-dimension spatial structure of the images in order to learn translation invariant features. These feature representations are particularly well-suited for fine-grained classification tasks line knee OA classification.

CNN Architectures for Knee Osteoarthritis Quantification

In [17], J. Antony *et. al.* quantify the severity of knee OA using different CNN models. The authors fine-tuned the **BVLC CaffeNet** [18] and **VGG-M-128** [19] pre-trained networks, modifying the top layer to train the model for classification and regression, leading to integer and real number label predictions, respectively.

Although they obtained good results (being the regression loss error lower than the classification one), the network architectures relied on *off-the-shelf* models that were previously trained on the ImageNet dataset with a high number of trainable parameters. In a posterior work [20], the authors presented a **customized CNN architecture** consisting of 5 convolutional layers (each followed by batch-normalization and ReLU activation layers), max-pooling layers and a fully connected layer (Figure 2.6). The network was trained from scratch for classification and for classification-regression all together. For the later, the network optimized a weighted-ratio of two loss functions, i.e. categorical cross-entropy (for classification) and mean squared error (for regression). This optimization provides the network with information about ordering of KL grades (penalizing high distance from the ground truth) and also with information about the quantization of the grades. The results obtained were better than the previous work, proving that multi-class classification and regression with the proposed methods provided a performance improvement.

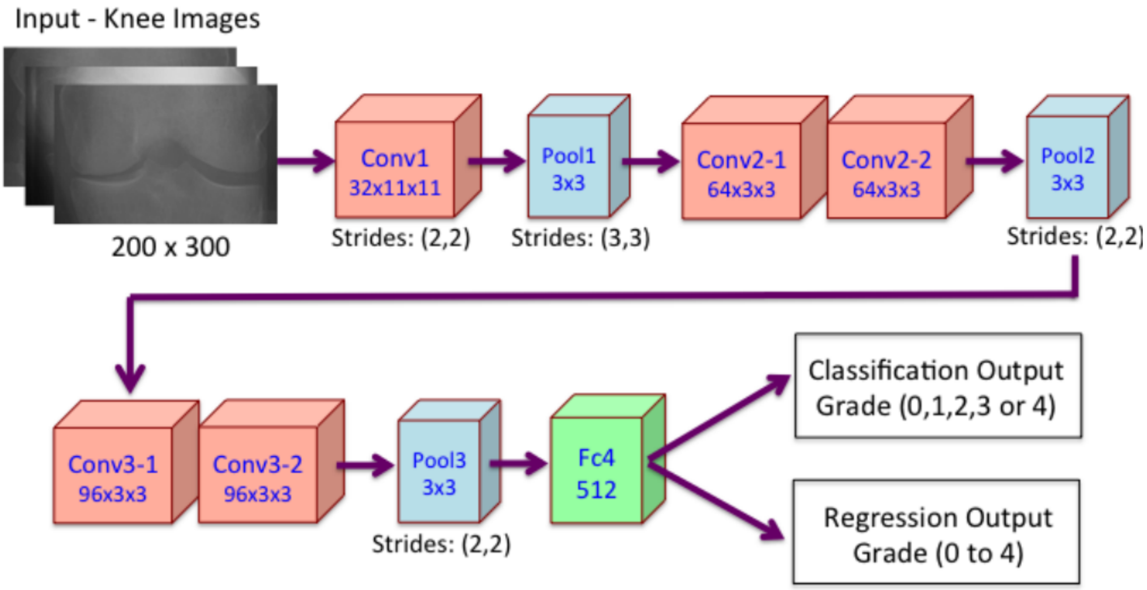


Figure 2.6: Network architecture for simultaneous classification and regression (extracted from [20])

Other authors [21, 22] followed the approach of using pre-trained models such as **DenseNet**

[23] to make OA assessments. This network consists of a series of convolutional blocks where each layer concatenates the feature maps of all preceding layers as input, keeping the output number of feature maps the same (Figure 2.7). This sequence of operations allows the network to learn from previous steps while keeping the number of learnable parameters low. Dense blocks make use of residual connections which were previously presented with ResNet architecture [24]. In contrast to ResNets, DenseNets do not combine features through summation but by concatenation as a means to transfer information further deep into the network layers. As stated in [23], DenseNet leads to better results than ResNet over ImageNet with the same number of learnable parameters, making this architecture an interesting option for knee OA classification.

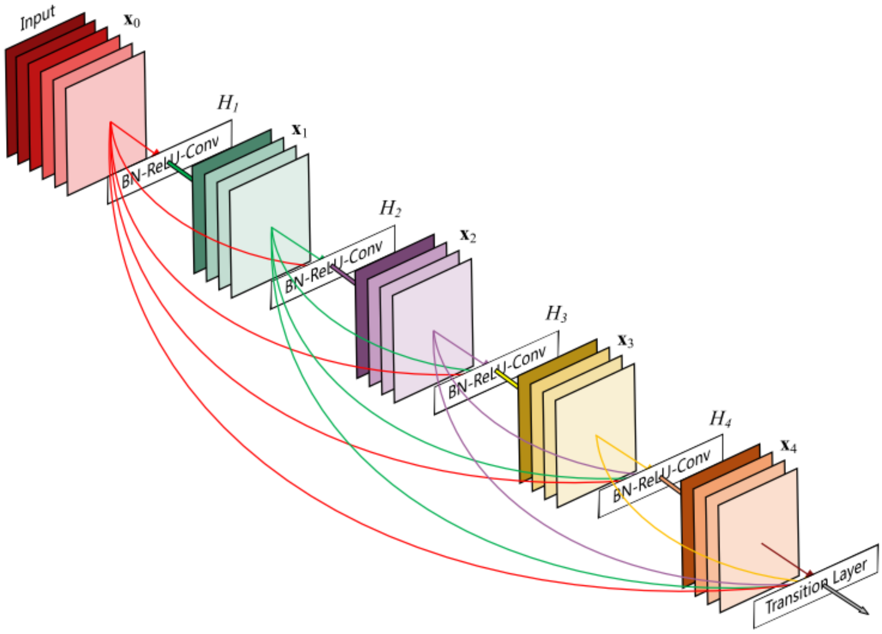


Figure 2.7: A 5-layer dense block of DenseNet with growth rate of k=4. Each layer takes all preceding feature maps as input (extracted from [23])

In [21, 22], the approach to knee OA classification is similar. The major difference is that [22] trained multiple models in order to ensemble the ones with better validation results. Among these models, the authors proposed variants which added demographic information to the network output, i.e. age, sex and race are concatenated to the flatten output from the DenseNet model.

All studies based on CNN presented above use X-ray images of the knees as input to the network with either no pre-defined orientation or keeping a side orientation for consistency (i.e. flipping one of right or left orientation). The later approach prevents the algorithm from having to learn an additional feature of side. A.Tiulpin *et. al.* [4] addressed this by inputting

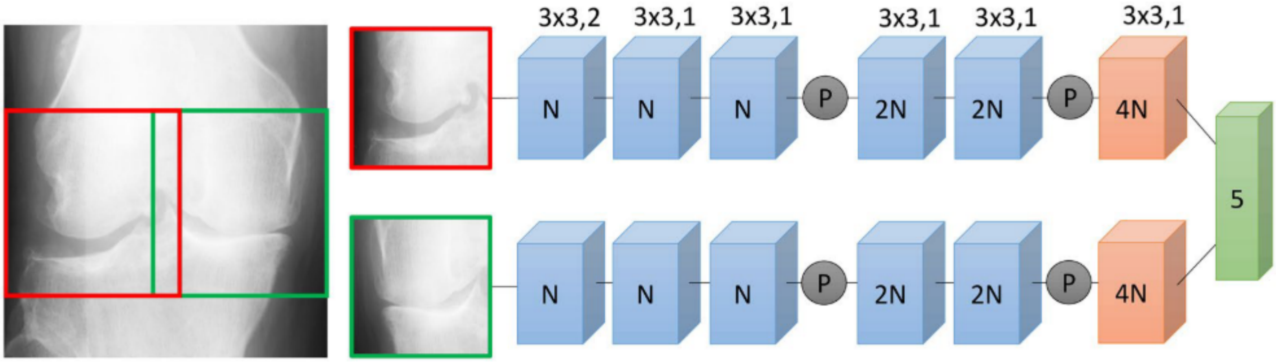


Figure 2.8: Schematic representation of the proposed Siamese network's architecture by [4]. First, knee joint sides are cropped (lateral on the left, medial on the right side) and flipped to match orientation. Then, the images are fed to each branch and the output is concatenated and fed to the last fully-connected layer before prediction (extracted from [4])

both medial and lateral sides of a knee joint separately. The authors leveraged the apparent symmetry of the knee joint by cropping medial and lateral sides of the knee and flipping the medial crop to match the orientation of the lateral side. Hence, the input images are oriented the same way prior to be fed to the network (Figure 2.8).

The other novelty of [4] relies on the utilization of **Deep Siamese CNN network architecture**. This architecture was originally intended for similarity comparison between pairs of images [25]. Traditionally, it consists of two branches with identical networks (shared weights) where each one corresponds to each of the input images. In this study, the authors modified the architecture in order to learn the same features from each knee side instead of comparison purposes. This approach follows the assumption that edge-detection features are not different for the lateral and medial side. Afterwards, the output of each branch is concatenated and fed to a fully-connected layer for the classification prediction. The branches share a custom network architecture consisting of convolutional blocks, max-pooling and a softmax layer.

Since this work will utilize a CNN-based model approach for knee OA assessment, all related works described in this chapter are shown in Table 2.1 for reference.

2.3.3 Automatic Detection of Knee Joints

The assessment of knee OA severity can be achieved by examining the radiographic features shown in a X-ray image, i.e. joint space narrowing, osteophytes formation and bone changes. The practitioner will search for knee OA features in particular regions of the joint; hence, it is essential to isolate the **Region of Interest** (ROI) to ease the computer-aided classification

Table 2.1: List of recent knee Osteoarthritis severity classification approaches using CNN-based models

Year	Authors	Methodology
2017	J. Antony *et. al.* [20]	Customized CNN architecture. Trained for classification and classification-regression.
2020	K. A. Thomas *et. al.* [21]	DenseNet pretrained network architecture. Trained for classification.
2018	B. Norman *et. al.* [22]	Ensemble of DenseNet pretrained network architecture. Trained for classification.
2018	A. Tiulpin *at. al.* [4]	Deep Siamese network consisting of branches of customized CNN network architectures. Trained for classification.

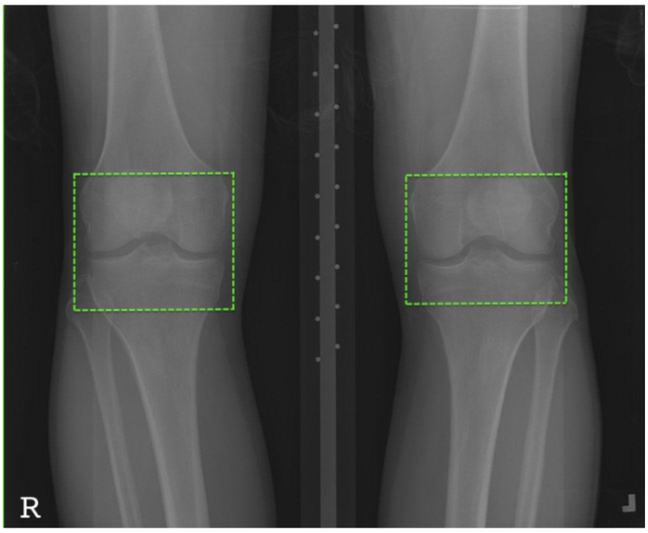


Figure 2.9: Knee OA X-ray with the region of interest (extracted from [20])

process. By extracting the ROI, non-relevant information in the joint surroundings will not be input to the network, it will make the image invariant to the position of the joint within the original X-ray, as well as it will reduce the image size. Hence, the computation cost and the model performance would be increased. Figure 2.9 shows a sample of a knee OA X-ray with the ROI.

There are several approaches for detecting and segmenting knee joints or specific parts. Their automation varies according to the manual intervention required, so developing fast and accurate ROI detection methodologies is essential for computer-aided diagnosis. Although challenging, **automatic ROI detection methods** are beneficial since they can be integrated to the classification pipeline, minimizing computation time, number of steps and human intervention.

L. Shamir *et. al.* [26] proposed a **template matching** approach in order to automatically detect the knee joint center. Template matching utilizes patches of selected knee joints with different KL grades and uses them as templates. Using a sliding windows over an image, the Euclidean distance between all patches and each window is calculated so that the window with the shortest distance is then recorded as the image center. After detecting the center, a segment of the image is extracted around it and utilized as ROI for the classification. A. Tiulpin *et. al.* [27] introduced a knee anatomy-based proposal that used **Histogram of Oriented Gradients** (HOG) as feature descriptors in order to detect the changes of image intensity due to the space between the bones. The location of the detected changes are used as region proposals and a Support Vector Machine (SVM) is trained to locate the ROI. Both approaches reported good results in knee joint ROI location.

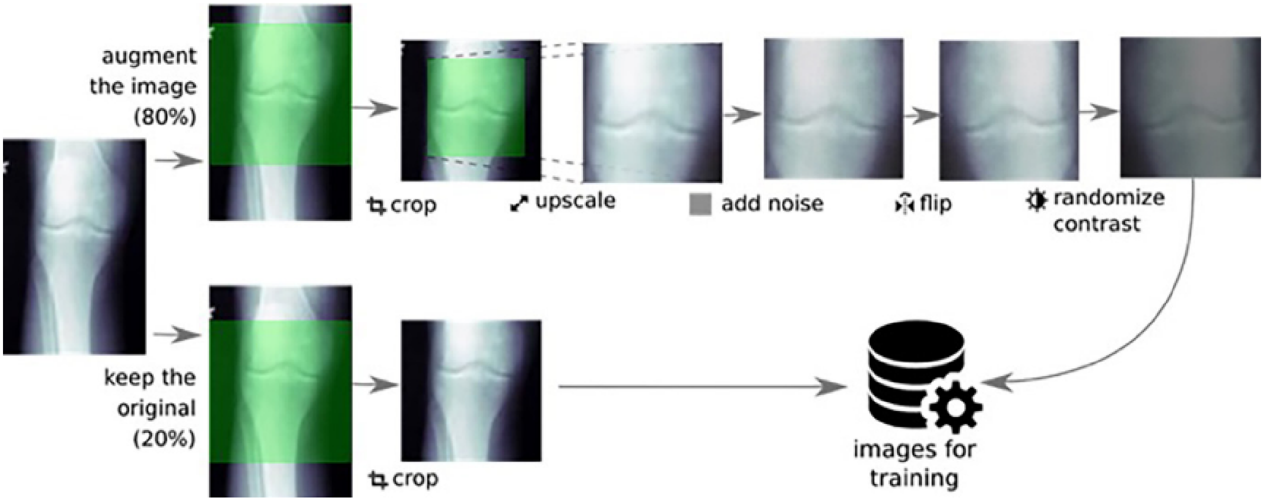


Figure 2.10: Knee OA X-ray with the region of interest (extracted from [20])

K. Thomas *et. al.* [21] followed a **CNN-based approach** for knee OA classification where the input images are not specific ROI as in other methodologies. In their study, the authors input either the original single-leg X-ray image cropped to a square image (by removing upper and lower rows) or an augmented version of it. The later is a subset of the cropped image where additional rows and columns have been removed in an uniformly distributed manner to keep 70% to 95% of the image (Figure 2.10). The authors justify this approach to create a model more robust to variability that, otherwise, would have been reduced with a cropped ROI.

J. Antony *et. al.* [20] proposed a methodology to automatically detect the ROI of the knee joint by using a **Fully Convolutional Network** (FCN). The authors input a training set of X-ray images with a binary mask that specifies the ROI as ground truth. The FCN architecture (Figure 2.11) consists of convolutional stages followed by max-pooling layers, and

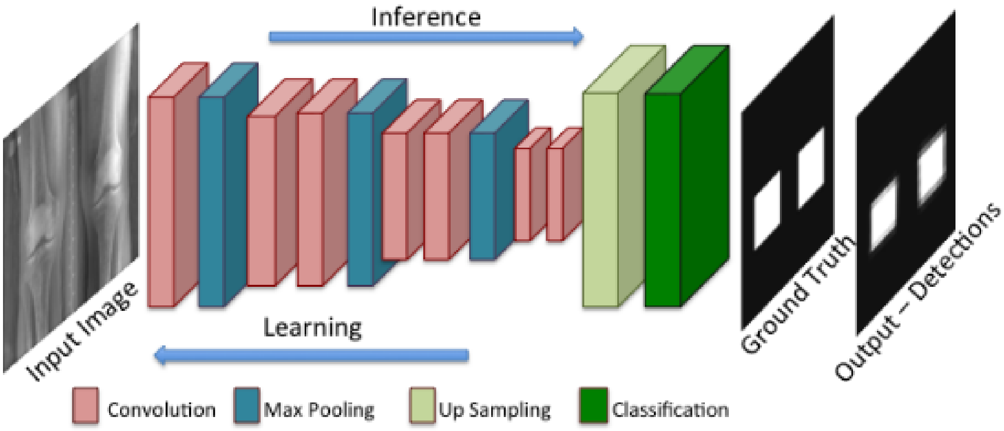


Figure 2.11: Knee OA X-ray with the region of interest (extracted from [20])

a final stage of up-sampling and a fully-convolutional layer with a sigmoid activation function for pixel classification. After the training, the bounding box is deduced by the contour of the detected ROIs and, after up-scaling to the original resolution, it is extracted for classification. According to the authors, FCN methodology is highly accurate with over 90% ROIs correctly detected.

P. Chen *et. al.* [28] followed a different CNN detection architecture, **YOLOv2** [29], considering the ROI detection as a regression problem. YOLOv2 architecture divides the input image in a grid and refines the height, width, center coordinates, and confidence score for each of the bounding boxes located. The grid-by-grid search approach is essentially the same as the sliding window strategy but the advantage of CNN here is that the features of all proposals are calculated all together in one forward operation. As result of an image forwarded through the model, different bounding boxes will be proposed. The bounding box with a higher confidence score is taken as the final detection result and re-scaled to the original image resolution (Figure 2.12).

2.3.4 Explainability of Neural Network Models

Deep Learning models automatically learn features from the images in order to produce the output. Learnt features may not be relevant to knee OA diagnosis as the models may react to background noise, image artifacts or other features that are not relevant to the disease. In order to obtain a qualitative understanding of how the model arrived at a particular prediction, it is helpful to visualize where the model is focusing its attention on the image.

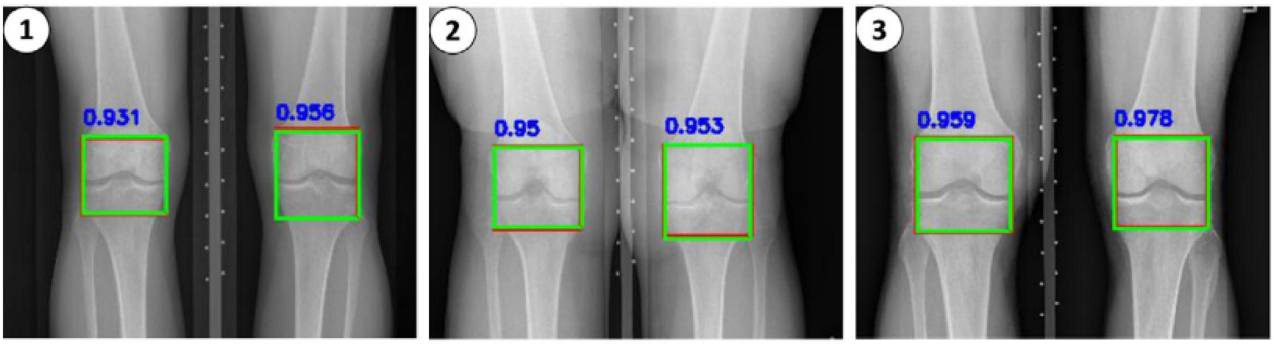


Figure 2.12: Region of interest for knees utilizing YOLOv2. Red bounding boxes correspond to annotated knee joint region, green bounding boxes stand for the detected knee joint region and blue scores are the intersection over union value between both boxes (extracted from [28])

Simonyan *et. al.* [30] proposed a methodology that relies on the back-propagation algorithm to compute the gradient of the logits with respect to the input of the network. The magnitude of the derivative indicates which pixels would need to be changed the least to affect the class score the most. Hence, the saliency map is a pixel-wise representation that indicates the relevance of a pixel for the class prediction, and such pixels would correspond to the relevant area locations in the image. With the gradient values, a visualization, commonly refer to as **Saliency Map**, can be obtain to highlight the class-relevant pixels. However, as shown in Figure 2.13, the saliency map is a discrete pixel representation and it is not class discriminative, i.e. relevant pixels may not correspond to the object but also other objects or background.

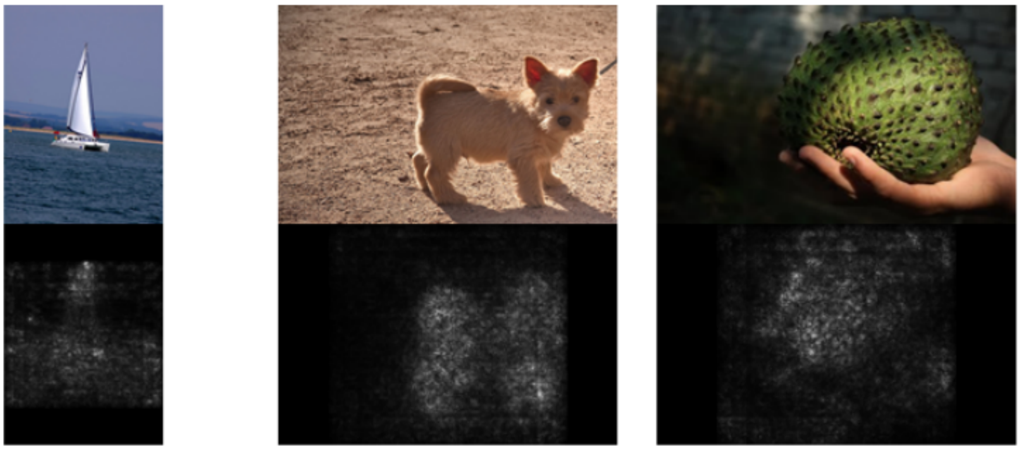


Figure 2.13: Example of saliency maps (bottom) for three images (top). Saliency map shows the class-relevant pixels (extracted from [30])

Alternative methodologies have been proposed based on **Attention Maps** in order to ease the category localization in the image. With this approach, the visualization typically consists of a heat-map type of image that can be superimposed over the original image. The highlighted areas would correspond to the class-specific discriminative regions, i.e. the section of the image where the network is focusing its attention to make a prediction. B. Zhou *et. al.* [31] introduced **Class Activation Mapping** (CAM) as an alternative explanation method for CNNs. CAM leverages the Global Average Pooling (GAP) layer that some networks architectures include prior to the last fully-connected layers. GAP averages the activations of each feature map (from the previous convolutional layer) and concatenates them in a vector of weights. By projecting this vector of weights on the convolutional feature maps, the most relevant features will be highlighted (Figure 2.14).

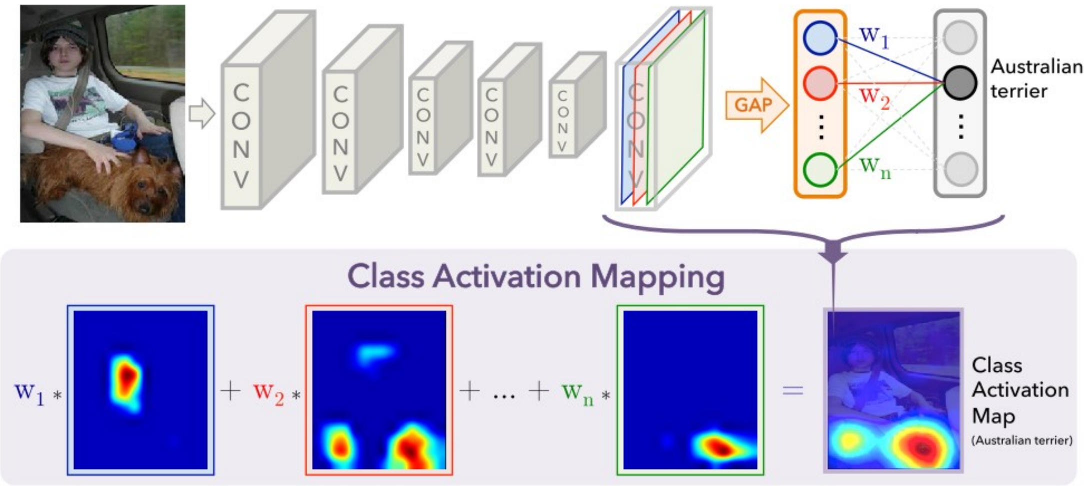


Figure 2.14: Class Activation Mapping (CAM). The CAM highlights the class-specific discriminative regions (extracted from [31])

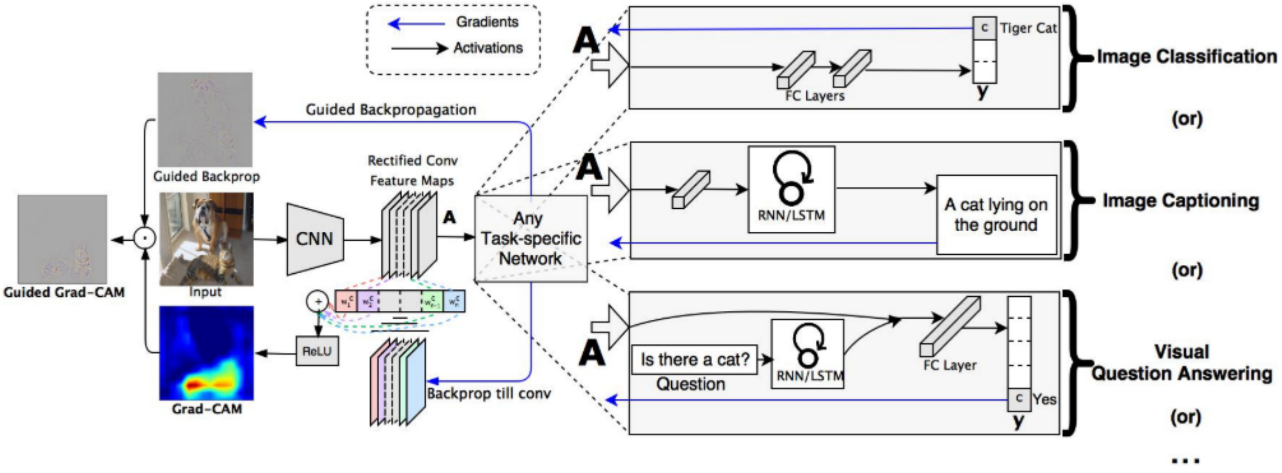


Figure 2.15: Grad-CAM overview (extracted from [32])

In order to be versatile and not rely on specific network architectures, R. Selvaraju *et. al.* [32] proposed **Gradient-weighted Class Activation Mapping** (Grad-CAM). To obtain the Grad-CAM heat-map, the gradient of the score for the class (before the Softmax activation layer) is computed with respect to the feature maps (Figure 2.15). Then, the gradients are global-average pooled over the dimensions of the feature maps to obtain the feature-importance weights; hence, Grad-CAM does not need a GAP layer to obtain weights. A weighted combination between the feature maps and the weights is then performed, resulting in a heat-map with the feature map size. After re-scaling to the original image size, the heat-map would highlight the class-specific discriminative regions.

2.3.5 Methodology Proposal

In this work, a methodology for automatic detection and quantification of knee OA severity will be proposed. This will be approached as an image classification problem, assessing the OA severity out of plain X-ray images. The process will involve different steps:

1. **Automatic detection and extraction of the ROI**. A CNN-based approach will be followed in order to detect the region of interest. The proposal consist on using the U-Net architecture [33] to segment the ROI out of X-ray images. This process includes the need for binary masks of the training and validation sets that will be generated for that purpose. Extraction of the ROIs will be carried out by generating bounding boxes around the contours of the detected knee joint areas.

2. **Data set definition for classification**. With the ability to detect and extract the ROIs, a specific data set including isolated knee joints will be created from the original data set. These images will define the input for the following steps.

3. **Quantification of knee OA severity**. A CNN-based approach will be followed in order to assess the knee OA severity. As baseline models, DenseNet [23], EfficientNet [34] and VGG [30] pre-trained architectures will be utilized as backbones for the classification process.

4. **Attention Map Visualization**. In order to evaluate the relevant features for the network to make a prediction, Grad-CAM methodology will be used to create a visual representation of the network attention.

Chapter 3

Methodology

3.1 Workflow overview

This chapter presents the approach to automatically detect and quantify knee OA severity from X-ray images. The methodology proposed in this work covers the entire pipeline from location of the knee joint to the knee OA severity assessment based on KL grade system. This process involves two main steps: 1) automatically detecting and extracting the ROI (by localizing the knee joint out of X-ray images), and 2) classifying the localized knee OA severity by assigning a KL grade.

Each of the steps will be described in depth in this chapter, including the introduction of the dataset used for the experiments, the CNN-based architectures that allow the detection of the knees and their OA severity classification, as well as the outcome of these methods and the conclusions. The code generated for the different sections can be found in a dedicated GitHub repository created for this study[1].

3.2 Dataset

The dataset used in this work was obtained from the controlled access datasets distributed from the Osteoarthritis Initiative (OAI), a data repository housed within the NIMH Data Archive (NDA). The OAI dataset consists of progression and incidence cohort subjects under the knee

[1] https://github.com/d-duran/KOA-location-diagnose-CNN

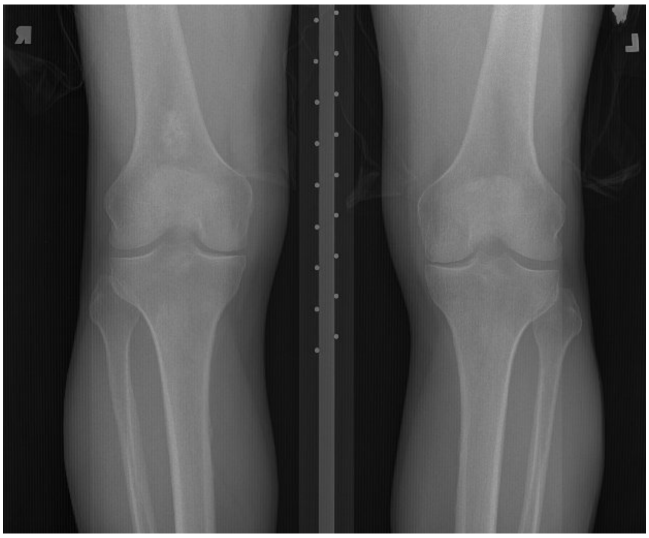


Figure 3.1: X-ray image from the OAI dataset

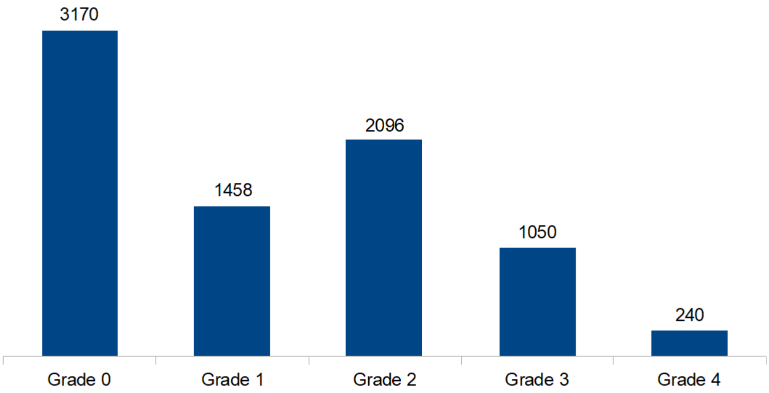


Figure 3.2: OAI dataset image distribution as per KL grades

OA study for which Kellgren & Lawrence (KL) grades have been assigned. Figure 3.1 shows a X-Ray image sample from the OAI dataset.

After an initial evaluation. a total of 4007 subjects X-Ray images from the baseline data was considered. Since each X-Ray image contains two knees and isolated knee images are required for this study, a total of 8014 knee X-Ray images were utilized. Figure 3.2 shows the distribution of the images as per KL grades.

3.3 Preprocessing

Since the dataset was collected as part of a progression OA study, the images vary in terms of contrast, brightness, etc. As a preprocessing step, a *contrast limited adaptive histogram equalization* (CLAHE) is performed in order to improve contrast and normalize the image levels.

In order to assess knee OA, medical practitioners examine only the knee joint region. Eliminating non-relevant information in the joint surroundings makes the image invariant to the position of the joint withing the original X-Ray image. Focusing on ROIs reduces the image size, hence the computation cost would be reduced and the model performance would be increased.

3.3.1 Automatic Detection and Extraction of Knee Joints

In order to generate a set of knee ROI images from the original dataset, an automatic detection of knee joints based on image segmentation is proposed. Image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. This process results in an image representation with multiple segments which allows for object location.

In this study, a CNN-based approach is followed in order to detect the ROI. The proposal consists on using the **U-Net** architecture [33] to segment the ROI out of the X-ray images. U-Net architecture (Figure 3.3) consists of two paths. First path is the contraction path, or **encoder**, which is used to encode the input image into feature representations at multiple different levels, capturing the context. The encoder is a stack of convolutional and max pool layers. The second path is a symmetric expanding path, or **decoder**, which semantically project the discriminative features learnt by the encoder onto the pixel space to get a dense classification. In order to get precise locations, skip connections are used at every step of the decoder in which the output of the previous step is concatenated to the feature maps from the encoder at the same level. As a result, the output is an image in which each pixel is classified to a particular class, i.e. *ROI* or *not ROI*; thus it is a pixel level image classification.

A subset of 200 X-ray images is extracted from the OAI datset. Each sample is then manually annotated by defining binary masks around the knee joints (Figure 3.4). The subset is randomly split into training (160 images) and test (40 images) sets. The ground truth for

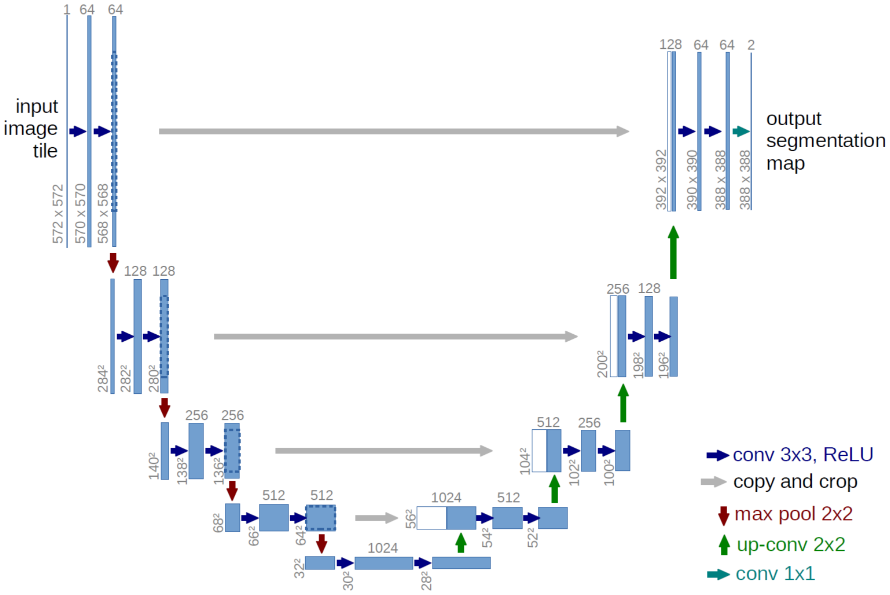


Figure 3.3: U-Net architecture representation [33]

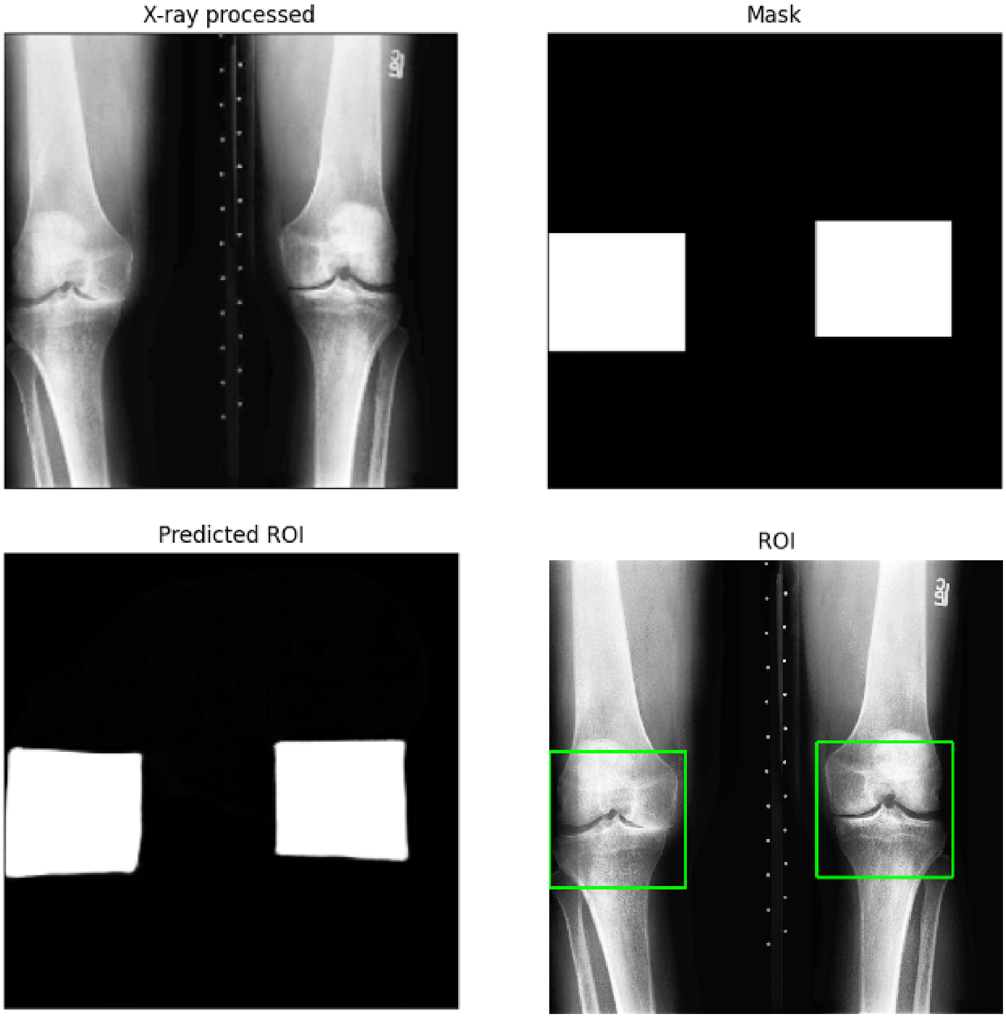


Figure 3.4: Preprocessed X-ray image (top-left), knee joint ROI binary mask (top-right), detected contour (bottom-left) and ROI bounding box (bottom-right)

training the network are the binary masks which delimit the knee joints ROI.

For this study, ResNet50 [24] off-the-shelf CNN network is considered to be the backbone of the U-Net. The network is pre-trained for general image classification on the ImageNet [18] dataset, which contains images of 1000 different classes. The use of this pre-trained network allows to leverage its already-learnt features such as edge detection, texture and patterns recognition for the encoding path, so that the networks does not need to be trained from scratch. For such purpose, *segmentation-models* library is used, which allow to chose among different backbone pre-trained architectures, including ResNet50.

After the training step is completed, the following steps are followed in order to extract the ROI:

- The output prediction of the U-Net is a binary image. Ideally, the areas corresponding to the knee joints (positive) will have 1 as pixels value and the remaining pixels will serve as background (negative), taking 0 as pixels value. The contours of the positive areas are located, as well as their centroids.
- A bounding box is created around the centroids such that the contour is fully contained. The bounding box is set to be square since this is the aspect ratio that forthcoming neural networks will expect as input.
- The bounding box is resized to the original X-ray image resolution and extracted as the ROI. Then, the image is resized to match the input size expected by the neural network.

Figure 3.5 shows a diagram with the complete process for detection and extraction of knee joint ROI.

3.3.2 Data preparation

In order to prevent the forthcoming algorithms from the need to learn additional feature of side, the ROI images from the left knees (located on the right side of the image) have been flipped vertically to match the right knees orientation. This procedure leverages the symmetry between the two knees in the original X-ray image. As a result, all extracted ROI images will present the medial and lateral side of the knees on the left and right side, respectively. Figure 3.6 shows a sample of a ROI extracted by the trained U-Net model.

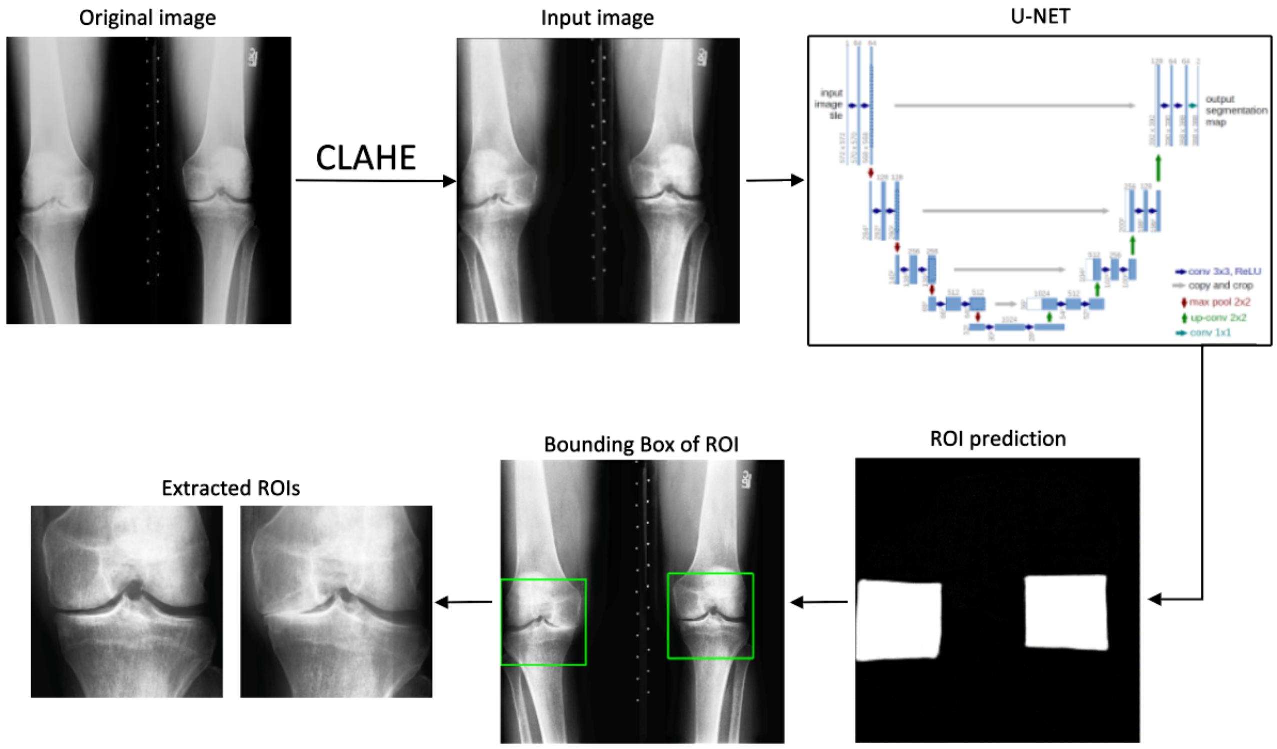


Figure 3.5: Detection and extraction of ROI pipeline

Knee OA radiographic features such as presence of osteophytes, joint space narrowing or bone deformity are relatively ease to identify to practitioners for late stages of the disease. Regarding neural networks accuracy, previous studies introduced in section 2 report lower misclassification rates for late stages (KL grades 2 to 4) in comparison to those corresponding to early stages of the disease, i.e. KL 0 and 1. Reported results prove that early stages of the disease are challenging for neural networks and special attention to those grades is required.

As stated previously, images of the dataset are imbalance as per KL grades. In order to ease the forthcoming neural networks learning, data augmentation is applied to the ROI dataset. Particularly, KL grades 0 and 1 are required to be as balanced as possible to reduce misclassification at early stages which are the most challenging to predict. The augmentations performed are as follows: the images are rotated a maximum of 20 degrees with a probability of 50%, and the brightness, contrast and gamma factor are randomly modified with a probability of 50%. An example of augmented images is shown in Figure 3.6. The data augmentation technique results in a new distribution as shown in Figure 3.7.

After the extraction and preprocessing of the knee joints ROIs, the new dataset needed to be organized in such a way that each image is assigned to a KL grade. All images belonging to the same class (KL grade 0 to 4) were automatically stored following the same directory. This process was carried out with the help of an auxiliary *CSV* file created for this purpose which contains the X-ray image filename as well as the KL grades.

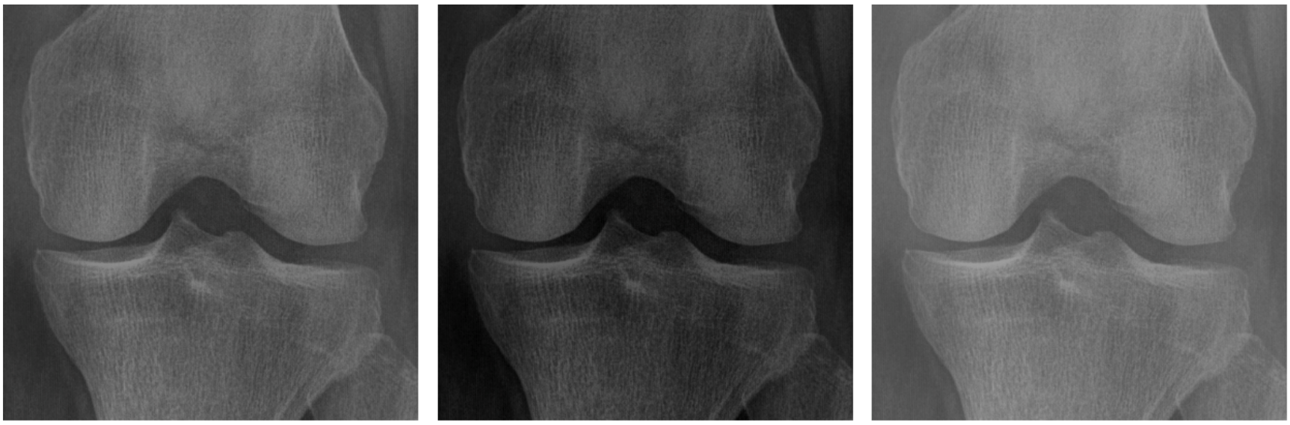


Figure 3.6: Knee joint ROI (left) with two augmentations (center and right)

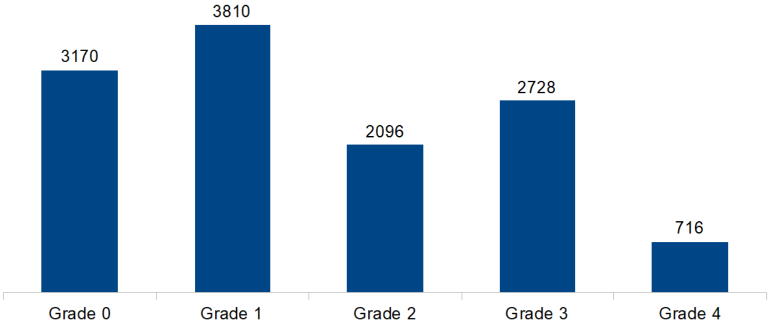


Figure 3.7: OAI dataset image distribution as per KL grades after data augmentation

3.4 Network architectures for Classification

Despite of the application of data augmentation technique, there are still classes under-represented in the dataset as depicted in Figure 3.7. To mitigate this issue, the associated weight to each class was calculated by using the *compute class weights* method from the Sklearn's *class weight* library. By using these weigths, the batches fed to the classifier will not over-sample any particular category.

In order to create the batches, Keras' *ImageDataGenerator* class is utilized. At this point, the required image preprocessing has already been applied, and the images size matches the input size expected by the neural network, i.e. 224 by 224 pixels. As common practice, the input images need to be normalized prior to be input to the neural network. This process includes rescaling pixel values from the range 0-255 to 0-1. The Keras' *preprocessing function* of the architectures considered is used for this step.

Additionaly, the *validation split* parameter is set to 0.2 to define the size of the validation set. The *ImageDataGenerator* method *flow from directory* is used to specify the path of the training and validation sets. The batch size is set to 16 and the *shuffle* parameter is set to *False* for the validation set since, later on, class labels need to be access unshuffled for evaluation. The image generator will then create batches that contains 16 preprocessed images and their corresponding class labels, represented with a one-hot encoded vector (e.g. [1,0,0,0,0] for KL grade 0).

For this study, off-the-shelf CNN architectures are considered to classify knee OA severity. These networks are pre-trained for general image classification on the ImageNet [18] dataset, which contains images of 1000 different classes. The pre-trained networks were fine-tuned for knee OA image classification based on the transfer learning approach. In transfer learning, the network is trained on external data (i.e. ImageNet dataset), and then the layers' weights are transferred to the target network. The lower layers of the network contain more generic features such as edge or texture, whilst the upper layers progressively focus on more task specific features. The top layers of the target network correspond to the final classification step, and are customize so that the output matches this problem characteristics, i.e. 5 target classes instead of 1000. During training on the new dataset, the layer weights of the target network will be slightly tuned based on performance.

Three different CNN architectures are considered in order to keep the model that yielded the best results possible. First, **VGG19** [30] network architecture is considered. In VGG19, the image is passed through a stack of convolutional layers where the filters use a very small receptive field (3×3), which is the smallest size to capture the notion of left/right, up/down, center. Since the model has over 144 million parameters, transfer learning is applied by using a model pre-trained on the ImageNet [18] dataset in order not to train the model from scratch. The layers placed after the convolutional blocks of the pre-trained model are replaced by a set of layers that matches the output size of this problem, i.e. a *Dense* layer with *ReLU* activation followed by a *Drop-out* layer, and a final *Dense* layer with 5 neurons and *Softmax* activation function.

Next, the **EfficientNet** [34] family of neural networks is considered. The authors obtained the baseline model, EfficientNet-B0, by means of balancing width, depth and resolution dimensions of the network. The dimensions search lead to a baseline model that allowed minimizing convolutional operations while increasing accuracy. The EfficientNet family of networks (B0 to B7) was constructed from the baseline model by scaling up the dimensions using the *compound coefficient*, ϕ. Among the EfficientNet variants, **EfficientNet-B4** was proven to yield the best results for this problem and, therefore, that will be the architecture configuration used herein. As described previously, the top layers that replace the original classifier are the same than those used in VGG19.

Inspired by the good results achieved by K.A. Thomas [21] and B. Norman [22], the **DenseNet** [23] architecture is also chosen to learn and extract the images features. This network consists of a series of convolutional blocks where each layer concatenates the feature maps of all preceding layers as input making use of residual connections. This architecture allows the network to learn from previous steps while reducing the vanishing gradient problem that deep networks usually present. Among the different implementations of the DenseNet architecture, Keras' **DenseNet121** is utilized due to its lower number of trainable parameters (8 million). In order not to train the model from scratch, transfer learning is applied by using a model pre-trained on the ImageNet [18] dataset. The top layers that replace the original classifier are the same than those used in VGG19 and described above.

3.5 Results evaluation

3.5.1 Automatic Detection of Knee Joints

Automatic detection of knee joints is carried out by solving an image segmentation problem. As described in section 3.3.1, an U-Net architecture based on a ResNet50 backbone is utilized for ROI detection.

The ground-truth images are binary masks corresponding to ROIs (positive regions) and background (negative regions). In terms of number of pixels, the predominant class is the background, hence, there is a class imbalance which may lead to unrealistic accuracy results.

The U-Net network is trained to minimise the *binary cross entropy Jaccard loss* function. The motivation behind the combination of binary cross entropy loss and the Jaccard loss is that the later computes the similarity between the ground truth and the prediction in a pixel-wise manner by means of the Jaccard index. Hence, the potential model's tendency to predict the predominant class to maximize accuracy is avoided.

The detection performance of the U-Net model is evaluated with the *Jaccard Index*. As described above, this metric measures the similarity between two sample sets evaluating the intersection over the union. In this context, the Jaccard index is calculated to compare all pixel values of the mask (ground truth) against the model's prediction, hence quantifying the ROI overlap. Jaccard index varies from 0 to 1, hence the higher the value the closer the prediction matches the mask ROI. A representation of the concept of intersection over union can be seen in Figure 3.8.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Where A and B are the mask and the prediction images.

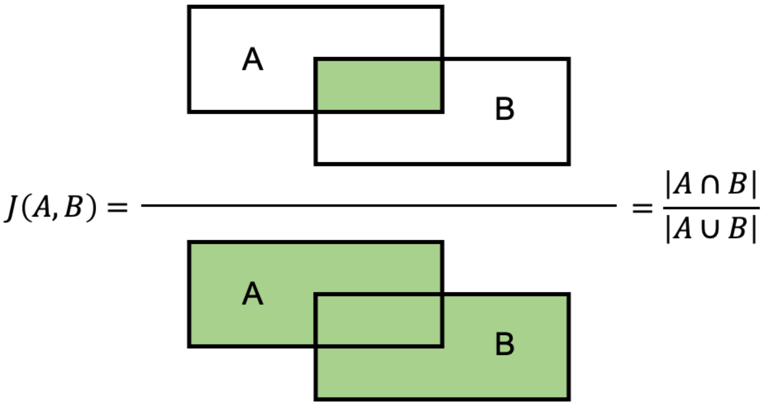


Figure 3.8: Jaccard index representation

3.5.2 Classification of Knee Osteoarthritis grade

3.5.2.1 Training loss and metrics

As mentioned before, the distribution of the examples across the dataset is not balanced. Evaluating neural networks performance in this scenario is not a trivial task since class imbalance might lead to slight bias during the network training.

Besides the data augmentation techniques applied to the dataset, it is possible to address the class imbalance problem by choosing a *loss function* that somehow considers the different available images per class. The loss function drives the optimization of the neural network weights, assessing the model performance during the training process.

In classification problems, the *categorical crossentropy* loss function (CE loss) is commonly used, which can be seen as the blue (top) curve in Figure 3.9. When using CE loss, even examples that are easily classified (i.e. $p \geq 0.5$) incur a loss with non-trivial magnitude. Over a large number of easy examples, the summed loss values can overwhelm the rare classes. This might be an issue when class imbalance is present in the problem as classes with more available images will be easier to classify for the network, eventually leading to high losses.

In order to address class imbalance, ***focal loss*** function [35] will be utilized to evaluate the classifier performance. Focal loss is an extension of CE loss that down-weight easy examples and focus training on hard negatives. Focal loss adds the modulating factor $(1 - p_t)^\gamma$ to the cross entropy loss, with *tunable focusing* parameter $\gamma \geq 0$. Focal loss is defined as:

$$FL(p_t) = -\alpha_t (1 - p_t)^\gamma log(p_t)$$

Where $\alpha_t \in [0, 1]$ is a weighting factor that may be set by inverse class frequency or treated as a hyperparameter, γ is the focusing parameter, and p_t is the model's estimated probability for the class with label $y = 1$, defined as:

$$p_t = \begin{cases} p & if y = 1 \\ 1 - p & otherwise \end{cases}$$

The γ parameter controls the shape of the curve (Figure 3.9). The higher the value of γ, the lower the loss for well-classified examples, so the attention of the model turns more towards examples which are hard to classify. Having higher γ extends the range in which an example receives low loss.

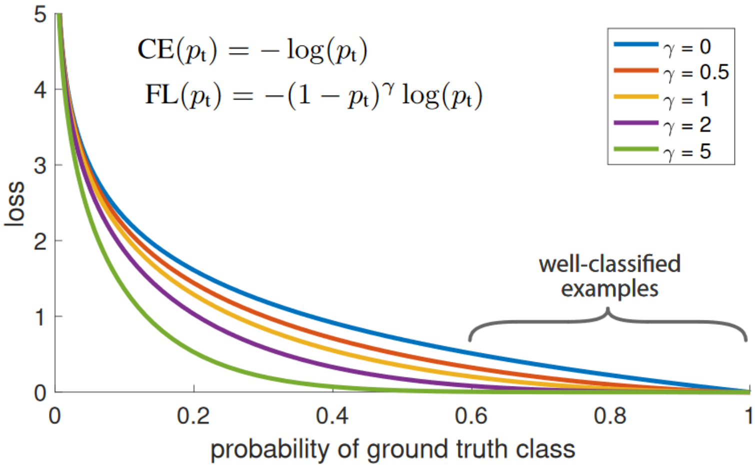


Figure 3.9: Focal loss visualization for different values of γ [35]

In order to assess whether the model obtains satisfactory results, different metrics can be used to evaluate its prediction capability. Since the network will predict a disease level, it is fundamental to ensure the agreement between the ground truth disease grade and the model prediction. In this regards, ***Cohen Kappa*** index measures the agreement between two raters who each classify samples into mutually exclusive categories. Cohen Kappa index is a quantitative measure of two raters agreement corrected for how often the raters may agree by chance.

Cohen Kappa score varies between -1 and 1, depending on the extent of agreement or disagreement. A score of 0 means there is random agreement between raters, whereas a score of 1 means that there is complete agreement. Therefore, a score lower than 0 means that there is less agreement that random chance.

Cohen Kappa score is calculated over a contingency table. First, the agreement by chance is computed (the sum of the products of the k marginal probabilities), and then this chance agreement is subtracted from the total observed agreement (the sum of the diagonal probabilities) before estimating the normalized agreement beyond chance.

$$k_w = \frac{\sum_{i=1}^{k} p_{ij} - \sum_{i=1}^{k} p_{i+}p_{+j}}{1 - \sum_{i=1}^{k} p_{i+}p_{+j}}$$

Where k is the number of classes, p_{ij} represents the proportion of all cases that receive the rating of i from the first rater and the rating j from the second rater, p_{i+} is the marginal distribution of the first rater's ratings, and p_{+j} is the marginal distribution of the second rater's ratings

In this particular problem, the classes are ranked as they correspond to different levels of knee OA disease, i.e. KL grades 0 to 4. This sense of order is important if the model mislabel a sample. For example, a model prediction of KL grade 0 with the ground truth being grade 4 should be considered differently than when ground truth and prediction are closer. This can be addressed by introducing weights to the Cohen Kappa score calculation that penalize wrong classifications based on the distance between the true value and the prediction, i.e. the higher the disagreement the higher the weight.

$$k_w = \frac{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij}p_{ij} - \sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij}p_{i+}p_{+j}}{1 - \sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij}p_{i+}p_{+j}}$$

Where w_{ij} are the weights associated with disagreement.

In this study, Tensorflow's ***Weighted Cohen Kappa*** metric implementation is used to calculate the model's performance. The weightage is chosen to be *quadratic*, hence penalizing the disagreement quadratically the higher it gets.

3.5.2.2 Explainability of results

Besides the performance metrics that can be obtained from a neural network model, the inherent complexity of architectures makes the results hard to interpret. In this work, a visualization methodology is followed in order to explain where the model focus its attention to obtain a categorical prediction.

As introduced in section 2.3.4, Grad-CAM methodology uses the gradients flowing from the logits (before *softmax* layer) into the last convolutional layer to assign importance values for a particular decision of interest. In order to obtain the class-discriminative localization, the gradient of the score y^c for a particular class c (e.g. KL grade 0) is computed with respect to feature map activations A^k, of a convolutional layer, i.e. $\frac{\mathrm{d}y^c}{\mathrm{d}A^c}$. The gradients flowing back are global-average pooled over the width and height (indexed as i and j, respectively) for each activation map to obtain the neuron importance weight, α_k^c.

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\mathrm{d}y^c}{\mathrm{d}A_{ij}^k}$$

A weighted combination is then performed between each weight-feature map pair in order to obtain a heatmap. As a result, the heatmap will indicate the areas whose intensity should be increased in order to increase y_c. Therefore, these areas correspond to the locations where the model focus its attention to detect class c.

The resulting heatmap size matches that of the last convolutional layer of the CNN; hence, it needs to be resized to the input image size in order to be superimposed for visualization.

Chapter 4

Experimentation

4.1 Workflow overview

This chapter presents the approach followed in order to 1) train the model for knee ROI detection and 2) to optimize the CNN-based architecture described in sections 3.3.1 and 3.4, respectively. The optimization process involves a set of experiments which include different CNN architectures, hyperparameters and training control techniques to find the best model.

4.2 Description of experiments

4.2.1 Automatic Detection of Knee Joints Experiments

As described in section 3.3.1, an U-Net architecture based on a ResNet50 pre-trained model is utilized for ROI detection.

The training process consists on inputting X-Ray images which are encoded by the network into feature representations at multiple different levels. Afterwards, the network decodes them in a symmetric expanding path and connects every level to its corresponding encoder's path level. The output consists of a pixel-wise binary image which contains the ROI and background regions. Prediction is then compared to the ground truth mask, evaluating the overlap with the Intersection over Union metric score as described in section 3.5.1.

The optimizer and hyperparameters set for training are chosen based on values commonly utilized in literature [20]. The optimizer is *Adam* with learning rates ranging from 0.001 to 0.0001, and exponential decay ranging from 0.8 to 0.99. The network is trained for 50 epochs and the batch size is 8. Figure 3.4 shows an instance of the test input, the ground truth and the output prediction of the U-Net.

4.2.2 Classification of Knee Osteoarthritis Grade Experiments

As described in section 3.4, three different CNN architectures have been considered for analysis. Since pre-trained networks are fine-tuned for knee OA image classification, a training strategy that involve the transferred weights to the target network must be defined.

During the training process, the value of the transferred weights can be frozen so that only those weights of the top classifier layers are learned and updated. Other option is to unfreeze the transferred weights and let the whole network adjust to the new dataset. Following this approach, the weights will slightly change, specially on the higher layers, in order to learn problem-specific features.

Initial experimentation showed that freezing the transferred weights while only updating the top layers ones conditioned the accuracy of the results. The approach of unfreezing the transferred weights proved to be unstable at the beginning of the training for different hyperparameters configurations.

In order to make the training stable and lead to the best results possible, the following alternative regarding the weights update is proposed:

- First, only the top layers weights are considered trainable during 3 epochs, hence the transferred weights are frozen. This will make the top layers to be trained towards the problem-specific characteristics.
- After the first 3 epochs, all layers weights are unfrozen and become trainable. This will make the whole network to be trained in forthcoming epochs, including the backbone.

Regarding the duration of training, the first stage is set to 3 epochs based on experience as it was observed that, after the third epoch, the validation loss and metrics kept stable and did not improve over time. The optimizer is Adam with a fixed learning rate of 0.001.

The second stage duration is set to 50 epochs, with the following callbacks consideration:

- Early stopping: training finishes when the validation loss does not improve over the last 3 epochs.
- Learning rate scheduler: learning rate is reduced by a factor of 0.1 when, for a given epoch, there has been no progress in terms of the validation loss for the last 2 epochs.

The optimizer is Adam with the learning rate value being defined as per the scheduler above. The initial learning rate value is given by the hyperparameter search as described in section 4.3.

4.3 Hyperparameter optimization

For knee OA classification, transfer learning approach has been followed for three different network architectures. The networks overall performance might be considerably impacted by the hyperparameters involved in the training process: learning loss getting stuck in local minimum due to high/low learning rates, long training time due to a large number of learnable parameters (number of neurons of Dense layers), etc.

In order to find the best hyperparameters possible, different techniques can be followed such as *Grid search* and *Random search*. The former defines a grid of hyperparameters values and the algorithm exhaustively searches this space in a sequential manner for every possible combination. This approach can be very inefficient in computer power and time as the number of hyperparameters increase. The later, instead of providing an explicit set of possible values, it provides a statistical distribution (e.g. exponential, normal, random,...) for each hyperparameter from which values are sampled.

These approaches perform individual experiments by building multiple models with various hyperparameters values. However, experiments are not able to use the performance information from the other experiments to be improved. In order to make use of this information, ***Bayesian hyperparameter optimization*** approach is used. Bayesian optimization is a sequential model-based optimization algorithm that uses the results from the previous iteration to decide the next hyperparameter value candidates. In order to do that, the method defines a probabilistic surrogate model to describe the objective function (loss or metric to be minimized/maximized) mapping hyperparameters to a probability of a score on the objective function. Based on this surrogate model, which updates in an iterative process, the algorithm balances the exploration and exploitation of the hyperparameters space to find the best combination.

In this work, the *BayesianOptimization* class from *bayes_opt* library is utilized to find the hyperparameters combination that leads to the best performance of the CNN architectures studied. The hyperparameters considered for optimization are the learning rate, the number of neurons in the top Dense layer and the dropout rate. The ranges of possible values have been chosen based on performance and the values commonly used in the literature [20, 22, 27]:

- Learning rate range: [0.0001 - 0.001]
- Number of neurons range: [1024 - 2048]
- Dropout rate range: [0.2 - 0.5]

The remaining hyperparameter of *batch size* involved in the training process have been set to 16 based on performance and computation capability.

Chapter 5

Results analysis

5.1 Automatic Detection of Knee Joints

The image segmentation model is trained using the U-Net architecture following the procedure described in section 3.3.1. Figure 5.1 shows the best performing model, corresponding to a learning rate of 0.0001 and exponential decay of 0.8. As shown in the figure, the learning curves converging to small loss when training this model.

Jaccard index (JI) is used to quantify the automatic detection of knee joint ROI. Intuitively, it can be interpret as the percentage of overlap between the ROI defined on the mask and that of the prediction. Per reference, $JI \geq 0.8$ means that the predicted ROI overlaps with 80% of the ROI defined on the mask. ROI masks are defined in a way that surrounding pixels are not part of the knee joint, hence a prediction with 80% of overlap would still include the joint and considered valid.

Table 5.1 shows the percentage of validation samples with knee joints correctly detected based on the Jaccard index values. As shown, the model is highly accurate with all of the samples overlaping at least 80% of the ROI area ($JI \geq 0.8$), and more than 2/3 of the samples overlaping at least 90% ($JI \geq 0.9$).

The high detection accuracy obtained with this model sets the base for the ROI detection and extraction pipeline followed in order to create the image dataset utilized for classification.

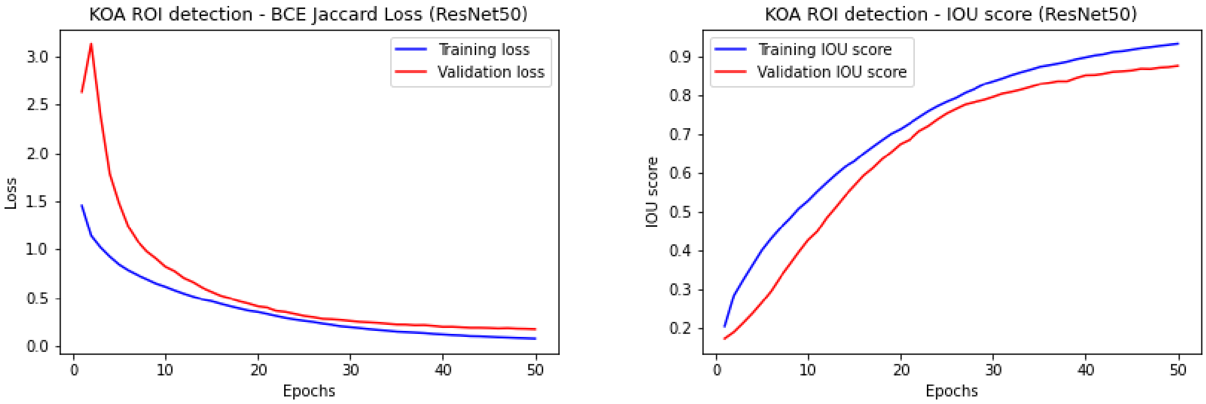


Figure 5.1: Evolution of BCE Jaccard loss (left) and Jaccard score (right) during U-Net model training.

Table 5.1: Evaluation of U-Net detection based on Jaccard Index (JI). Number of validation images which predicted ROI overlaps at least 50%, 80% and 90% with the ROI mask.

Validation Data	$JI \geq 0.5$	$JI \geq 0.8$	$JI \geq 0.9$
40 images	40/40	40/40	31/40

5.2 Classification of Osteoarthritis KL Grade

The resulting hyperparameter search carried out on the three CNN architectures and the metrics obtained can be seen in Table 5.2. This table shows the models hyperparameters which led to the best results for categorical crossentropy and focal loss functions. Loss and metrics are evaluated for the validation set.

From Table 5.2 it can be observed that EfficientNet and DenseNet architectures results are similar and higher to those of VGG. Based on the Cohen kappa score, EfficientNet-B4 has slightly better performance than DenseNet121. Despite the quantitative measure of similarity

Table 5.2: Hyperparameters obtained after Bayesian Optimization for VGG19, EfficientNet-B4 and DenseNet121 architectures. Results shown for categoriacal crossentropy and categorical focal loss functions evaluated on the validation set.

Loss function	Architecture	dropout	learning rate	neurons	loss	cohen kappa	avg f1-score	weighted f1-score
categorical crossentropy	VGG19	0.33	0.0008	1228	1.013	0.75	0.55	0.55
	EfficientNet-B4	0.45	0.0003	1210	0.699	**0.88**	0.70	0.68
	DenseNet121	0.38	0.0002	1183	0.702	0.86	0.72	0.70
categorical focal	VGG19	0.45	0.0003	1210	0.321	0.82	0.68	0.66
	EfficientNet-B4	0.41	0.0001	2017	0.261	0.87	0.69	0.68
	DenseNet121	0.33	0.0003	1650	0.248	**0.87**	**0.75**	**0.72**

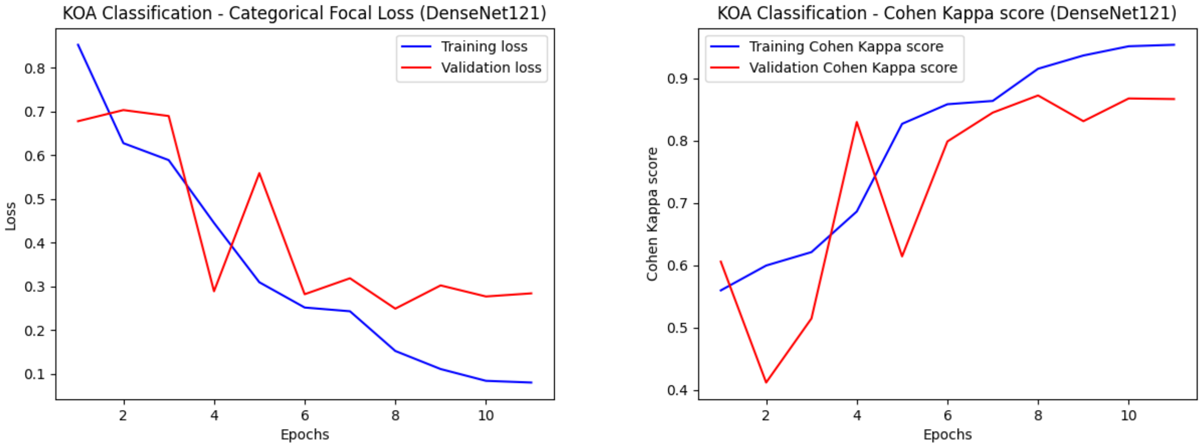


Figure 5.2: Evolution of loss (left) and Cohen kappa score (right) during DenseNet121 model training.

showed by the kappa score, the *weighted F1-score* has been also obtained to understand the models performance per class. In this regards, DenseNet121 shows the best results with a weighted f1-score of 0.72. Based on the results from Table 5.2, **DenseNet121** is considered as the best model due to its metrics scores.

Figure 5.2 shows the evolution of the categorical focal loss and Cohen kappa metric for the training and validation sets. During the training process, the model seems to converge to an optimal solution in less than 10 epochs. The overfitting is avoided thanks to the early stopping condition set to the training.

The resulting metrics for the model, including specific values per class, are shown in Table 5.3 whilst Table 5.4 shows the confusion matrix for the multi-class classification. The model achieves an accuracy score of 72% on the validation set, as well as 72% of f1-score weighted average. This metrics show a good performance of the model in the classification task.

It can be noted that classification of ranked grades is challenging, particularly in KL grades 0 and 1, where misclassification is higher compared to the other grades. As previously introduced, variation of OA disease features such as joint space and osteophyte formation are hardly distinguishable in the early phases of the disease and it and can be easily misclassified as similarly reported by other studies.

Nevertheless, the results show that the network is able to learn a good representation of the different knee OA disease features resulting in high accuracy classification performance and low mislabeling rates as shown in Table 5.4.

Table 5.3: Metrics of DenseNet121 with categorical focal loss function for classification

Class	Precision	Recall	f1-score
0	0.71	0.73	0.73
1	0.71	0.59	0.64
2	0.56	0.73	0.64
3	0.91	0.79	0.85
4	0.79	0.99	0.88
accuracy			0.72
macro avg	0.74	0.77	0.75
weighted avg	0.73	0.72	0.72
cohen kappa			**0.87**

Table 5.4: Confusion matrix for multi-class classification using DenseNet121 model

		Predicted class				
		0	1	2	3	4
True class	0	**480**	110	36	0	0
	1	170	**460**	140	15	0
	2	31	57	**310**	25	0
	3	0	14	63	**430**	35
	4	0	0	0	1	**130**

In order to contextualize the results, Table 5.5 compares the results obtained in this work with recent state-of-the-art studies. The table collects different metrics reported by the authors, including precision, recall and f1-score, as well as average scores. Note that some metrics values (such as weighted average values) needed to be obtained from their reported confusion matrices.

As shown in Table 5.5, results obtained by our DenseNet121 outperform the other authors results. Particularly, our DenseNet121 performs a better job in per-class metrics, with higher weighted average precision, recall and f1-score.

A particular characteristic on B. Norman *et. al.* [22] work is that KL grades 0 and 1 are considered as a single disease grade. The authors made this choice per recommendation of their internal clinical radiologist since the clinical response for these two grades are usually the same. Under this assumption, our DenseNet121 also presents better results in global and per-class metrics.

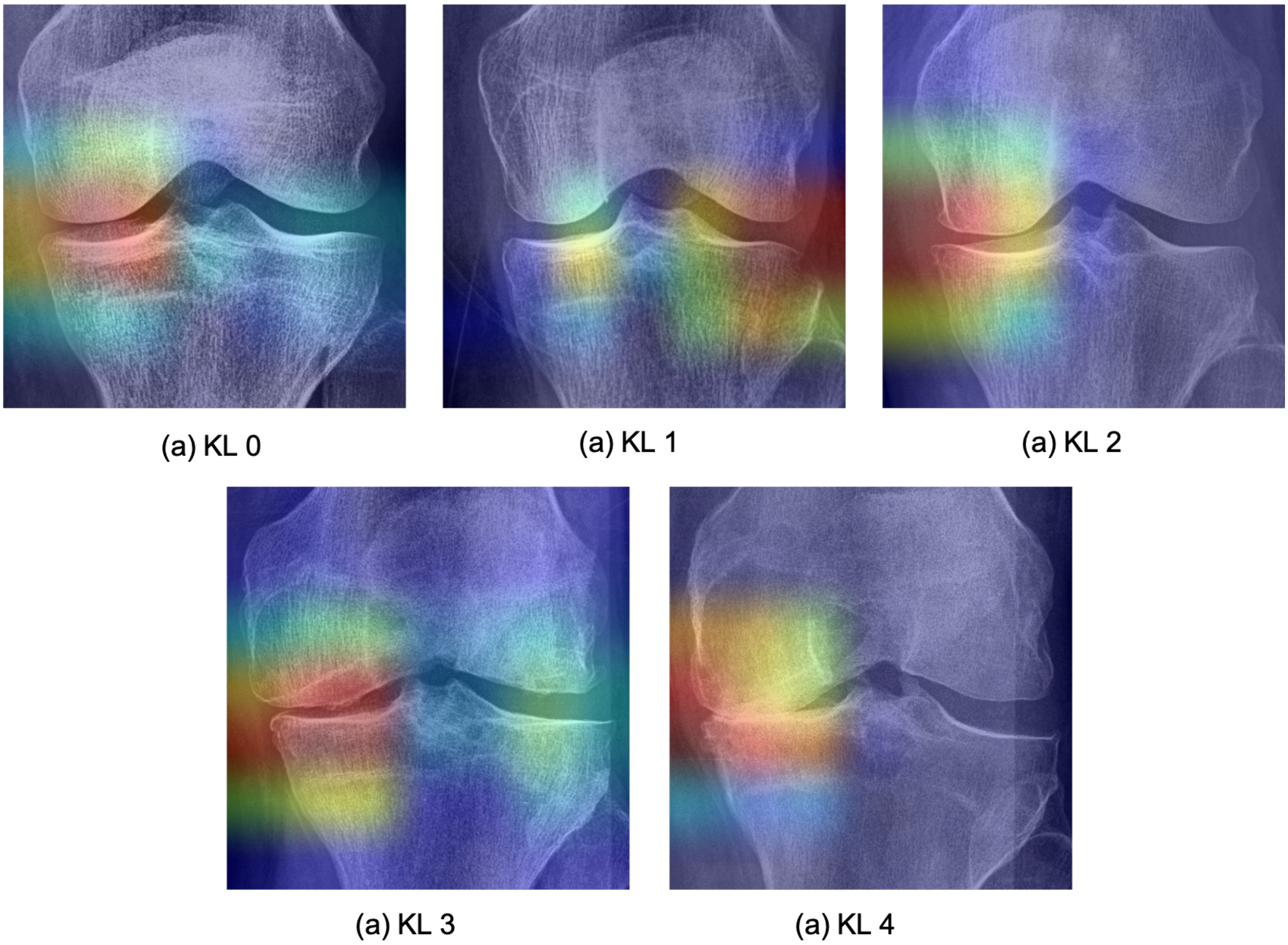
Table 5.5: Classification of knee OA grade metrics for DenseNet121. Results are compared to collected or calculated data from reported results.

	Ours			J. Antony [20]			K. Thomas [21]			B. Norman [22]		
Class	Precision	Recall	f1-score	Precision	Recall	f1-score	Precision	Recall	f1-score	Precisoin	Recall	f1-score
0	0.71	0.77	0.74	0.68	0.80	0.74	0.73	0.87	0.79	0.89*	0.84*	0.86*
1	0.72	0.59	0.65	0.32	0.15	0.20	0.38	0.27	0.31			
2	0.56	0.73	0.63	0.53	0.63	0.58	0.71	0.67	0.69	0.58	0.70	0.63
3	0.91	0.79	0.85	0.78	0.74	0.76	0.82	0.81	0.81	0.80	0.69	0.77
4	0.79	0.99	0.88	0.81	0.75	0.78	0.87	0.86	0.87	0.81	0.86	0.83
accuracy	-	-	**0.72**	-	-	0.64 (2797/4400)	-	-	0.71 (2890/4090)	-	-	0.78 (4217/5381)
macro avg	**0.74**	**0.77**	**0.75**	0.62	0.61	0.61	0.70	0.69	0.70	0.77	0.77	0.77
weighted avg	**0.73**	**0.72**	**0.72**	0.60	0.63	0.61	0.69	0.71	0.69	0.80	0.78	0.79
cohen kappa	-	-	**0.87**	-	-	-	-	-	-	-	-	-

*The authors considered KL grades 0 and 1 as a single class.

5.3 Explainability of classification

Grad-CAM visualizations were obtained as described in section 3.5.2.2 for the trained model and input images. Figure 5.3 shows an example of Grad-CAM visualization for each KL grade. The figure shows that the strongest network activation comes from the medial and lateral joint margins as well as the area between the joint bones, providing the highest contribution to the model's prediction. The most probable reason is that the model learnt knee OA radiological features such as presence of osteophytes and joint space narrowing in order to assess the grade of the disease.



(a) KL 0

(a) KL 1

(a) KL 2

(a) KL 3

(a) KL 4

Figure 5.3: Visualization of attention maps of classified examples for each KL grade.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This Master's thesis focuses on robust methods used for localization of knee joints out of a X-ray image and prediction of knee Osteoarthritis diagnosis. The main goal is to develop computer aided diagnostic tools to assist professionals on the assessment of knee Osteoarthritis (OA) severity by developing deep learning based automatic methods.

A summary of the investigations, research findings, experimental results, and the proposed solutions in this thesis is as follows.

Chapter 1 introduced knee OA degenerative joint disease, the diagnostic features and grading index. The motivations for this thesis are described, the research objectives are presented, as well as the project development steps that set this thesis structure.

Chapter 2 deeper describes knee OA disease and its prevalence, incidence and economical impact, concluding that the burden of knee OA to society is high and that the implications of knee OA are not to be underestimated. In this chapter the literature in knee OA assessment is reviewed, including clinical imaging utilized, the computer aided approaches for diagnostics of knee OA, as well as the necessary technical and terminology background are introduced.

Chapter 3 presented the approach to automatically localise and extract knee joints in radiographs. An image segmentation approach was implemented for the automatic localisation of knee joints. The segmentation is based on binary masks as ground truth which set the region of interest where the knee joint is located within the X-ray image.

Chapter 3 also presented the approach to classify the knee OA severity grade from the localized knee joints. Pre-processing of the images, including the data augmentation techniques applied to address the class imbalance problem are described. The pre-trained CNN architectures considered for the classification task are introduced. Based on the available knee OA grade labels, the metrics and loss functions considered to evaluate the models performance are presented. As part of the model performance evaluation, the visualization methodology used to explain where the model focus its attention to obtain a prediction of disease severity is described.

In Chapter 4, different CNN architectures were tested, including the architecture trained for the image segmentation of the knee joints ROI, as well as the architectures for knee OA severity classification. The hyperparameters optimization methodology followed to select the optimal architecture and hyperparameters is described.

In Chapter 5 the outcome of the experiments are shown for the localization of knee joints and for the knee OA severity classification. A significant outcome of this thesis is a methodology that achieves high accuracy in localizing knee joints out of plain X-ray images, as well as a very good performance in the OA diagnostic assessment of KL grades, outperforming recent studies in macro and per-class metrics.

The following is a brief summary of the research contributions of this thesis.

- Proposing a highly accurate technique to automatically detect and localise the knee joints from the X-ray images by means of a **U-Net neural network architecture**, considering the task as a segmentation problem.
- Developing a classifier based on pre-trained convolutional neural networks (CNN) to assess knee OA severity that is **highly accurate and outperforms recent studies**.
- Proposing an approach to train a CNN with a **focal loss function** to account for the inherent problem of class imbalance.
- Implementing a **Bayesian hyperparameter optimization** to account for previous experiments in the CNN hyperparameter search.
- Evaluating the performance of the models with **Cohen Kappa index** to measure the agreement between the predictions and the ground truth and account for misclassification according to the rank of KL grades.

- Developing a **Grad-CAM** visualization methodology to explain where the model focus its attention to obtain a prediction of disease severity.

Table 6.1 summarizes the accomplishment of the objectives that were defined in section 1.3. All of the objectives are considered as accomplished through the complete development of the study and experiments that have been carried out.

Table 6.1: Summary of objectives accomplishment

Objective	KPI	Accomplished	Comments
Design and implementation of a CNN architecture to automatically detect and extract the ROI of a knee X-ray image.	Functional scripts capable of generating cropped images of the ROI.	Yes	The model is highly accurate with all of the samples overlapping at least 80% of the ROI area ($JI \geq 0.8$), and 2/3 of the samples overlapping at least 90% ($JI \geq 0.9$).
Design and implementation of a CNN architecture to assign a OA severity grade to a knee X-ray image.	Functional scripts capable of generating a prediction label of the localized ROI.	Yes	The model achieves a Cohen kappa score of 0.87, an accuracy score of 72% on the validation set, as well as 72% of f1-score weighted average.
Evaluation and interpretation of the classification output against the ground truth.	Selection and justification of the best metric, and assessment of the model performance.	Yes	Jaccard index utilized to measure the similarity between the ground truth mask and the segmented prediction evaluating the intersection over the union. Weighted Cohen kappa index utilized to measure the agreement between the ground truth and class prediction.
Define a methodology to obtain the class-discriminating activation maps in order to examine the region where the network is detecting the disease-relevant features.	Being able to plot the activation maps obtained.	Yes	Grad-CAM visualizations show that the strongest network activation comes from the medial and lateral joint margins as well as the area between the joint bones.

6.2 Future Work

During the development of this study, new lines of work have been identified that relate to improving the performance of the knee OA severity assessment as well as to serving as the foundation for a supporting diagnosis tool.

The main goal is to develop computer aided diagnosis tools that assist practitioners in assessing knee OA severity. This study has set the foundation to that purpose by defining a workflow and a methodology that allow to localize and extract a knee joint from X-ray images, and to perform a prediction of disease severity. Further steps would consist of combining all individual steps into a single pipeline, i.e. image pre-processing, knee location and extraction, disease severity prediction and visualization. The resulting workflow can be contained into an API hosted on a cloud-based website. This would allow non-technical users to benefit from the model's prediction capability by uploading a single image.

As regards the network architectures considered in this study, the usage of pre-trained networks on the ImageNet dataset proved to be beneficial in terms of the results achieved when transfer learning was applied. Nevertheless, there are other alternatives that would be interesting to be explored for the classification task such as training a custom CNN from scratch or adopting a different type of architecture, i.e. Vision Transformers. The former would require a time-consuming optimization process as well as longer computation time. The later would consider the images as a sequence where network's attention and memory are the foundation for feature learning.

Since the OAI image dataset utilized in this study is part of a progression knee OA study, there are available multiple images for each pacient at different time periods. With the achieved ability of disease severity prediction, it would be interesting to consider the problem as a time-dependent sequence. Therefore, exploring the progression of the disease by analyzing the X-ray images would lead to a risk assessment. Here, different deep learning based approaches could be follow such as RNN or LSTM architectures, although Visition Transformers have recently be proven as a potential better candidate when sequential inputs are considered.

As it was introduced in section 2, knee OA is characterized by cartilage degradation and bone change, hence early detection of OA is relevant since early treatment could prevent cartilage and bone loss. Since evidences suggest that changes in bone occur early in the development of OA, it is interesting to explore a methodology that allow for early detection of OA by focusing the analysis on bone structure.

Early phases of knee OA are related to KL grades 0 and 1. According to the results reported in section 5, these KL grades are particularly challenging to predict since variation of OA disease features are hardly distinguishable in the early phases of the disease.

Some studies in the literature have explore the analysis of bone texture to get insights from the specific bone patterns at different disease phases. Inspired by this studies, a new line of investigation has been followed after the completion of this thesis regarding the analysis of subcondral bone texture. This process involves the following steps:

1. Defining a methodology to locate keypoints in the knee joints that allow to locate the region of interest. This methodology consists of creating a set of ground truth keypoints located at the left and right sides of the tibia, and each of the two intercondylar eminence (spines) on the center area (see Figure 6.1).

2. Training a neural network, i.e. MobileNet-V2, to be able to automatically locate the keypoints.

3. Locating the trabecular bone region of interest at the medial side of the knee. The ROI is defined as a square with its right side on the vertical of the medial intercondylar eminence and the top side low enough to avoid the subchondral cortical plate (top edge of the tibia) where sclerosis appear (see Figure 6.1).

4. Extracting the bone ROI and pre-process the image to enhance the bone texture (see Figure 6.1).

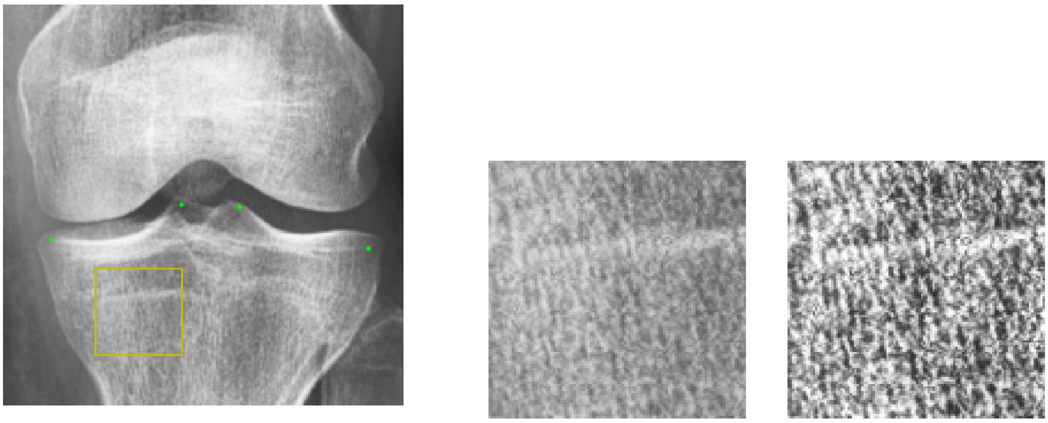


Figure 6.1: Keypoints of knee joint for bone ROI location (left), extracted bone ROI (center) and pre-processed bone ROI (right)

Despite this investigation is not yet been concluded, it is worth sharing that results show bone texture analysis is promising for disease severity prediction when utilized as input for a CNN neural network. Further investigation is required in regards to pre-processing of the bone ROI as well as the possibility of combining this approach with this thesis study. The assessment of knee OA severity might be improved by the combination of different features, i.e. radiological features such as presence of osteophytes and joint space narrowing; and the analysis of bone texture. Particularly, it is interesting to study the effect of this new source of information in regards to the early stages of the disease (often framed within KL grades 0 and 1) in order to reduce the misclassification for lower grades and provide early detection capability.

Glossary

backbone Feature extracting network which is used within the neural network architecture. This feature extractor is used to encode the network's input into a certain feature representation..

cartilage Cartilage, or cartilaginous tissue, is a resilient and smooth elastic tissue, rubber-like padding that covers and protects the ends of long bones at the joints and nerves..

osteoarthritis Osteoarthritis is a degenerative joint disease that affects both the cartilage and the bone and soft tissues of the joint. It is part of rheumatic diseases and, within this classification, it is a type of arthritis..

osteophyte Bony lumps that grow around the knee joint.

sclerosis Slow-growing lesions to the bone that happen very gradually over time..

subchondral bone Bone lying immediately beneath the articular cartilage.

subchondral bone plate Subchondral bone plate is a thin cortical lamella, lying immediately beneath the calcified cartilage in a joint..

subchondral trabecular bone Trabecuale are small tissue elements in the form of a small beam, strut or rod that supports or anchors a framework of parts within a body or organ. Subchondral trabecular bone refers to trabeculae in subchondral bone..

Bibliography

[1] F. V. Wilder, B. J. Hall, J. P. Barrett, and N. B. Lemrow, "History of acute knee injury and osteoarthritis of the knee: a prospective epidemiological assessment the clearwater osteoarthritis study," *Osteoarthritis and Cartilage*, vol. 10, pp. 611–616, 2002.

[2] M. C. Wick, M. Kastlunger, and R. J. Weiss, "Clinical imaging assessments of knee osteoarthritis in the elderly: A mini-review," *Gerontology*, vol. 60, pp. 386–394, 2014.

[3] J. H. Kellgren and J. S. Lawrence, "Radiological assessment of osteo-arthrosis," *Annals of the Rheumatic Diseases*, vol. 16, pp. 494–502, 12 1957.

[4] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, and S. Saarakkala, "Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach," *Scientific Reports 2018 8:1*, vol. 8, pp. 1–10, 1 2018.

[5] E. Ridge, "Introducing guerrilla analytics," *Guerrilla Analytics*, pp. 3–15, 1 2015.

[6] M. J. Lespasio, N. S. Piuzzi, E. Husni, M. . George, F. Muschler, A. J. Guarino, and M. A. Mont, "Knee osteoarthritis: A primer," *The Permanente Journal/Perm J*, vol. 21, pp. 16–183, 2017.

[7] G. S. Dulay, C. Cooper, and E. M. Dennison, "Knee pain, knee injury, knee osteoarthritis work," *Best Practice Research Clinical Rheumatology*, vol. 29, pp. 454–461, 6 2015.

[8] A. Cui, H. Li, D. Wang, J. Zhong, Y. Chen, and H. Lu, "Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies," *EClinicalMedicine*, vol. 29-30, p. 100587, 2020.

[9] M. Cross, E. Smith, D. Hoy, S. Nolte, I. Ackerman, M. Fransen, L. Bridgett, S. Williams, F. Guillemin, C. L. Hill, L. L. Laslett, G. Jones, F. Cicuttini, R. Osborne, T. Vos, R. Buchbinder, A. Woolf, and L. March, "The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study," *Annals of the Rheumatic Diseases*, vol. 73, pp. 1323–1330, 7 2014.

[10] J. H. Salmon, A. C. Rat, J. Sellam, M. Michel, J. P. Eschard, F. Guillemin, D. Jolly, and B. Fautrel, "Economic impact of lower-limb osteoarthritis worldwide: a systematic review of cost-of-illness studies," *Osteoarthritis and Cartilage*, vol. 24, pp. 1500–1508, 9 2016.

[11] A. J. Teichtahl, A. E. Wluka, M. L. Davies-Tuck, and F. M. Cicuttini, "Imaging of knee osteoarthritis," *Best Practice Research Clinical Rheumatology*, vol. 22, pp. 1061–1074, 12 2008.

[12] C. Buckland-Wright, "Subchondral bone changes in hand and knee osteoarthritis detected by radiography," *Osteoarthritis and Cartilage*, vol. 12, pp. 10–19, 1 2004.

[13] G. Li, J. Yin, J. Gao, T. S. Cheng, N. J. Pavlos, C. Zhang, and M. H. Zheng, "Subchondral bone in osteoarthritis: insight into risk factors and microstructural changes,"

[14] G. Stachowiak, M. Wolski, T. Woloszynski, and P. Podsiadlo, "Detection and prediction of osteoarthritis in knee and hand joints based on the x-ray image analysis," *Biosurface and Biotribology*, vol. 2, pp. 162–172, 12 2016.

[15] A. Brahim, R. Jennane, R. Riad, T. Janvier, L. Khedher, H. Toumi, and E. Lespessailles, "A decision support tool for early detection of knee osteoarthritis using x-ray imaging and machine learning: Data from the osteoarthritis initiative," *Computerized Medical Imaging and Graphics*, vol. 73, pp. 11–18, 4 2019.

[16] A. Brahim, R. Riad, and R. Jennane, "Knee osteoarthritis detection using power spectral density: Data from the osteoarthritis initiative," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11679 LNCS, pp. 480–487, 9 2019.

[17] J. Antony, K. Mcguinness, N. E. O'connor, and K. Moran, "Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks,"

[18] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding,"

[19] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets,"

[20] J. Antony, K. Mcguinness, K. Moran, and N. E. O'connor, "Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks,"

[21] K. A. Thomas, Łukasz Kidziński, E. Halilaj, S. L. Fleming, G. R. Venkataraman, E. H. G. Oei, G. E. Gold, and S. L. Delp, "Automated classification of radiographic knee osteoarthritis severity using deep neural networks content codes," *Radiology: Artificial Intelligence*, vol. 2, 2020.

[22] B. Norman, V. Pedoia, A. Noworolski, T. M. Link, and S. Majumdar, "Applying densely connected convolutional neural networks for staging osteoarthritis severity from plain radiographs," *Journal of Digital Imaging 2018 32:3*, vol. 32, pp. 471–477, 10 2018.

[23] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks,"

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition,"

[25] S. Chopra, R. Hadsell, and Y. Lecun, "Learning a similarity metric discriminatively, with application to face verification,"

[26] L. Shamir, S. M. Ling, W. W. Scott, A. Bos, N. O. Member, T. M. S. Member, D. M. Eckley, L. Ferrucci, I. G. G. Member, I. L. Shamir, N. Orlov, T. Macura, and I. Goldberg, "Knee x-ray image analysis method for automated detection of osteoarthritis,"

[27] A. Tiulpin, J. Thevenot, E. Rahtu, and S. Saarakkala, "A novel method for automatic localization of joint area on knee plain radiographs,"

[28] P. Chen, L. Gao, X. Shi, K. Allen, and L. Yang, "Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 84–92, 7 2019.

[29] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger,"

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[31] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization,"

[32] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization,"

[33] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation,"