FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING

Amir Mohammad Kazemtarghi

THE ROLE OF BONE MARROW LESIONS IN KNEE OSTEOARTHRITIS: TEXTURAL ANALYSIS OF SUBCHONDRAL BONE

Master's Thesis
Degree Programme in Computer Science and Engineering
July 2020

Kazemtarghi A. (2020) The Role of Bone Marrow Lesions in Knee Osteoarthritis: Textural Analysis of Subchondral Bone. University of Oulu, Degree Programme in Computer Science and Engineering, 59 p.

ABSTRACT

Osteoarthritis (OA) is the most common joint disorder in the world that affects various joints particularly hand, hip, and knee joint. The knee OA has been identified as the most impactful OA because it is the major cause of disability worldwide. Generally, OA progression leads to joint replacement surgery and causes enormous amount of financial costs. Thus, it is crucial to diagnose OA at an early stage and prevent or slow down its progression. Currently, clinical diagnosis of OA includes physical examination and clinical imaging. However, they are insensitive to early OA changes. On the other hand, it was shown that several imaging bio markers can be captured at an early stage of the disease. One of the important imaging bio markers for OA is the alternations of subchondral bone texture. Besides, there are other factors that cause these alternations such as bone marrow lesions (BML).

Two sub-studies have been conducted in this thesis. The aim of the first sub-study is to investigate the association between BML and OA diagnosis by using subchondral bone texture from plain radiography. OA subjects are defined by Kellgren-Lawrence (KL) grading scale. KL grade 0 and 1 represent no OA and grade 2, 3, and 4 are OA subjects. In this work, subjects at the baseline (first visit) of osteoarthritis initiative (OAI) dataset were selected. Then, they were categorised into three groups including subjects who has BML in medial tibia (group 1), subjects without BMLs at all (group 2), and lastly the subjects without medial tibia BMLs (group 3). In the next step, region of interest (ROI) was selected at the margin of medial tibia in plain radiographs. After that, 29 textural features from 4 textural descriptors including grey-level co-occurrence matrix (GLCM), histogram of image, absolute gradient, and fractal signature analysis (FSA) were computed from the extracted ROI. Subsequently, Fisher's exact test and Mann–Whitney U test were used in order to discover how textural features change among OA and non OA subjects in each group (first analysis) and how those differences change across the groups (second analysis).

Our results showed that there are significant textural differences between OA and non OA subjects when they have BMLs at medial tibia. Moreover, there were no significant textural differences among subjects with no BMLs and subjects

with no BMLs in medial tibia. These results indicate that the presence of BML as well as its location at subchondral bone may have association with OA incidence.

In the second sub-study, for research oriented purposes we built a deep convolutional neural network (CNN) based models to automatically detect OA from subchondral bone texture according to the Kellgren-Lawrence (KL) grading scale. We selected subjects without BMLs to make a fair comparison between magnetic resonance imaging (MRI) data and plain radiographs. In this study, subjects with no BMLs who have sagittal 3-D Double Echo Steady State sequence (3D DESS MRI) and plain radiography at the baseline of OAI were selected. In both imaging modalities, square sized ROIs were chosen located at the marginal region of medial tibia. Confusion matrix and area under the receiver operating characteristics curve (ROC AUC) were used to evaluate the model performance. Our results demonstrated that when subjects do not have BMLs, our model was not able to detect OA from the subchondral bone texture.

Keywords: osteoarthritis, subchondral bone texture, bone marrow lesions, medial tibia, region of interest, Fisherś exact test, Mann–Whitney U test, convolutional neural network

TABLE OF CONTENTS

ACKNOWLEDGMENT

We would like to thank OAI for proving high quality and well-organized images. The OAI is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2- 2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health. This manuscript was prepared using an OAI public use data set and does not necessarily reflect the opinions or views of the OAI investigators, the NIH, or the private funding partners.

Also, this work was supported by Research Unit of Medical Imaging, Physics and Technology (MIPT), Oulu, Finland.

Oulu, July 31st, 2020

Amir Mohammad Kazemtarghi

LIST OF ABBREVIATIONS AND SYMBOLS

OA	Osteoarthritis
OAI	The Osteoarthritis Initiative
MRI	Magnetic Resonance Imaging
AC	Articular Cartilage
MOAKS	MRI Osteoarthritis Knee Score
KL	Kellgren-Lawrence Semi-Quantitative Grading Scale
BML	Bone Marrow Lesion
ML	Machine Learning
CNN	Convolutions Neural Network
ROI	Region of Interest
GLCM	Grey-Level Co-occurrence Matrix
OR	Odds Ratio
FD	Fractal Dimension
FSA	Fractal Signature Analysis
SD	Standard Deviation
NPWNZG	Number of Pixels With Non-Zero Gradient
DESS	Double Echo Steady State

1. INTRODUCTION

1.1. Osteoarthritis

Osteoarthritis (OA) is a common joint disease and it is one of the leading cause of pain and disability throughout the world [1]. Based on recent OA studies and according to the United Nations statistics, it has been predicted to come up to even higher ranks in the foreseeable future [2]. The reason behind is that the prevalence of symptomatic OA is mostly associated with rising age due to depletion in regenerative capacity and muscle activity [2] (Figure 1). Today, there are more than 703 million people aged 65 or older which this number is projected to reach 1.5 billion by 2050 [1, 3]. Currently, millions of people (18% of women and 10% of men aged over 60 year) throughout the world are suffering from OA [4], which means that OA is connected to extensive financial costs in many countries. To illustrate, in 2016 the UK experienced 185,000 primary hip and knee replacements because of OA disease. This number is increasing rapidly because of ageing population [4]. In US, the annual rate of total knee replacement has increased by a factor of two since 2000 and the total number of non-surgical treatment process during each year now surpasses 640,000, with a total annual cost of about $10.2 billion [5]. New Zealand has reported the constant rise in the number of knee and hip replacement surgery during last decade [6] (see Figure 2).

OA may occur in various joints but the prevalence is higher for knees, hands, hips and spine [4]. The prevalence of knee OA is generally high indicated by multiple studies [7, 8, 9, 10, 11, 12]. For instance, in Framingham Study, age-standardised prevalence of radiographic knee OA was roughly 19% for subjects aged 45 and older [13]. In the Johnston County Osteoarthritis Project, the prevalence rate was 27% from the same age group and nearly 37% of subjects in the third National Health and Nutrition Examination Survey (NHANES III) were suffering from radiographic knee OA [13, 14]. Higher prevalence rate of knee OA means having more negative impact on individual life and socioeconomic status [15, 16]. From individual perspective, progression or incidence of OA leads to activity limitations, participation restriction alongside with negative effects on life quality, mood and sleep [16]. From socioeconomic perspective, some studies estimated that indirect costs of OA are eight times greater than direct costs of OA even though in most countries the direct costs are mainly considered [16, 17]. In conclusion, the burden of OA and its effects on individual productivity and socioeconomic costs are already substantial. Immediate actions should be taken to prevent or slow down the progression of this disease as our world is considerably growing older [16].

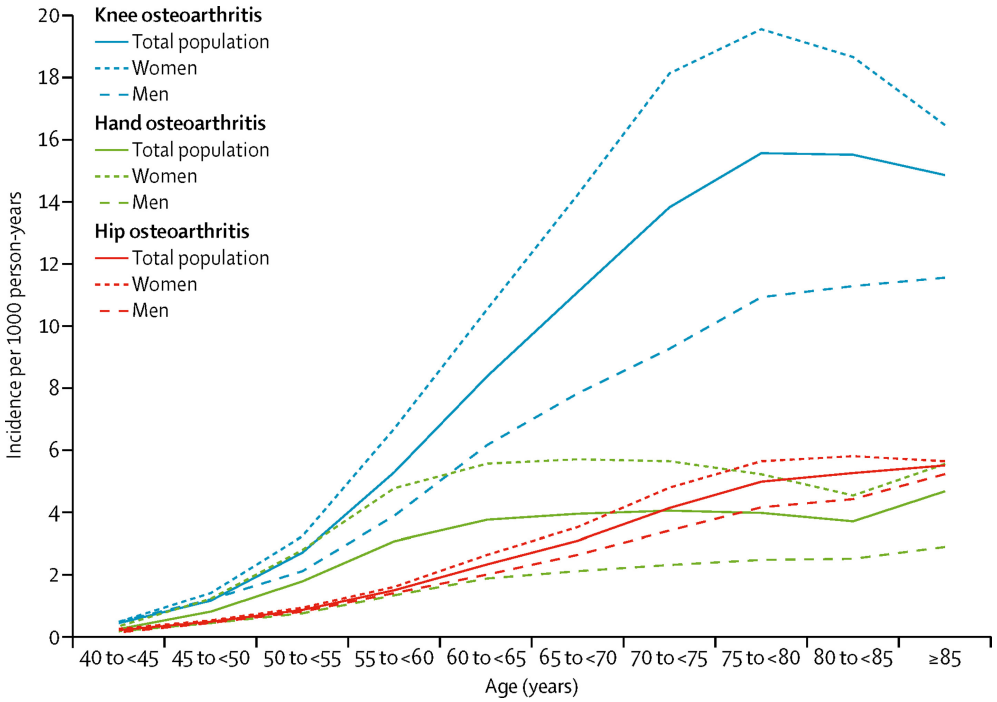


Figure 1. The reports of OA incidence from Spanish and UK general practice registry indicated that risk factor of age, between 50-year to 75-year old period, has more effect on knee joint rather than other joints and for female gender is much more male gender. These figures illustrates age-specific and gender-specific incidence (per 1000 person-years). This data is acquired from the randomly selected individuals from Catalonia (Spain) [18].

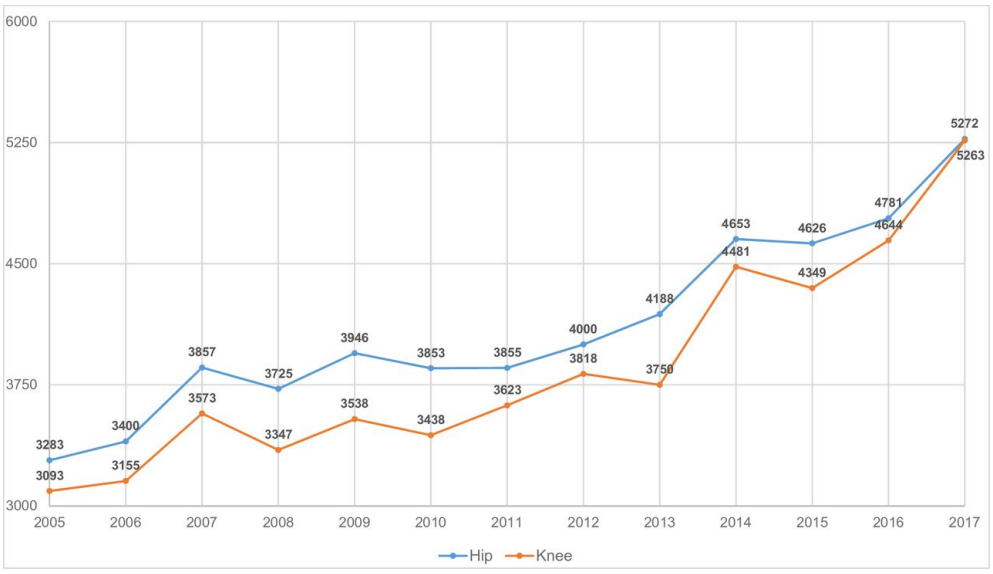


Figure 2. According to statistics from National Minimum Dataset (NMD) in New Zealand, Number of knee and hip replacements in New Zealand has steady risen from 2005 to 2017 by 70.2% and 60.6% respectively [6].

One of the most important characteristics of OA is being heterogeneous considering its onset and progression. For instance, knee OA might remain in the same stage over decades or progress rapidly in a few years [19]. Hence, recognising its mechanism and its risk factors is a crucial task for prevention of the disease and developing the treatment strategies. Enormous number of studies have been conducted to investigate risk factors of OA [20, 4]. According to recent study[21], OA risk factors may be mainly divided into two groups: 1) Systemic factors, such as age, genetics, body mass index, lifestyle, nutrition, sex hormones, bone density, 2) Mechanical factors including joint structure and malalignment, trauma, physical activity, muscle strength. It should be noted that not all of the aforementioned factors are used for OA diagnosis [22, 4].

Over a long period of time, it was believed that OA was mainly implies cartilage degeneration [23]. Currently, however, OA is perceived as a group of joint diseases or a multi-factorial disease [24]. For instance, structural changes in the various tissues including cartilage degeneration, damaging articular cartilage (AC) structure and function, remodelling of subchondral bone, emerging subchondral bone cysts, and marginal osteophytes formation, may simultaneously happen as OA progresses [25, 23, 24]. Recently it has been discovered that the cartilage is not the main source of pain in the early stage of OA [26, 27]. Instead, other factors such as remodelling in subchondral bone, synovial inflammation, damaging AC and ligament cause primarily the pain or inflammation [20] (Figure 3).

1.2. Subchondral Bone

Several subchondral bone definitions based on its morphology and mechanical properties can be found in literature [28, 29, 30, 31]. In this thesis, subchondral bone refers to the bony lamella (cortical end-plate) lying directly to the calcified zone of the articular cartilage which has two main components: 1) subchondral bone plate and 2) spongy trabecular bone. subchondral bone plate is Located under cartilage zone by cement line (see Figure 3). Its anatomical features such as thickness, density, and composition, vary from one subject to another [32]. Trabecular bone has more porous structure than subchondral bone plate [33]. As opposed to AC, subchondral bone tissues have blood vessels and nerves in the marrow space where AC can receive its nutrition from the blood supply [34]. Moreover, this structure allows subchondral bone to absorb mechanical stress transmitted through the joint and this can affect its structure as it has the capability of remodelling with mechanical load changes [35, 25]. According to the recent publication of MacKay et al. [36], it is estimated that subchondral bone structural changes may occur even before cartilage loss. Therefore,

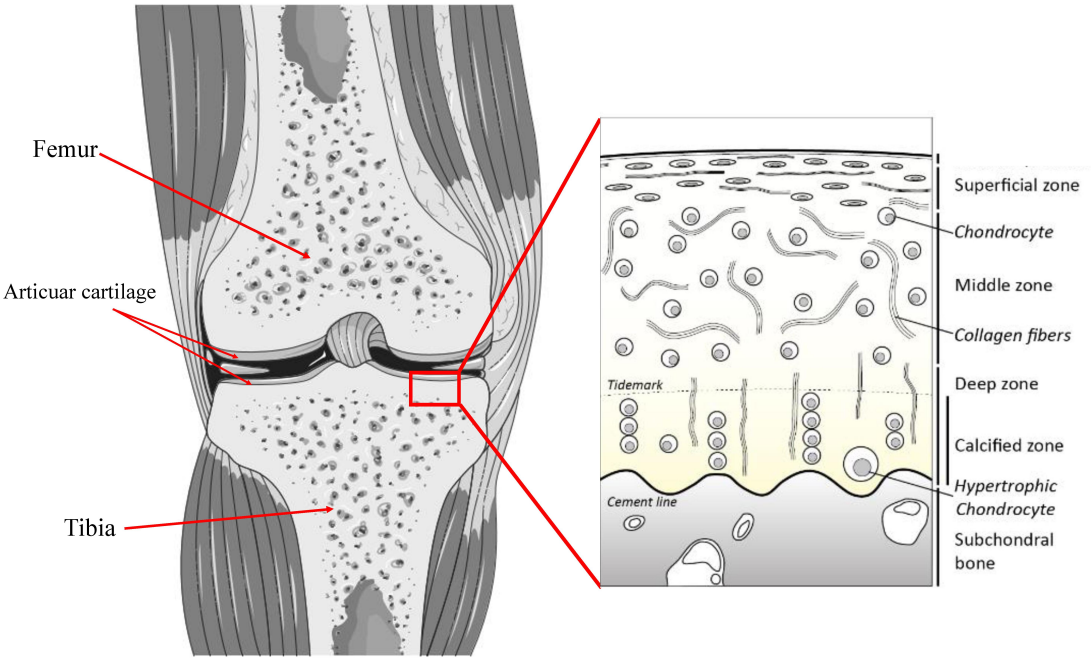


Figure 3. Schematic figure of a knee joint (left) and an articular cartilage (AC) (right). Different zones of AC are shown in the figure

subchondral bone has a potential as a biomarker in terms of knee OA detection in the early stages of knee OA.

1.3. Bone Marrow Lesions (BML)

Bone marrow lesions (BMLs) often occur when fluid builds up in the bone marrow in trabecular bone as a response to an injury or other conditions such as radiographic knee OA [37]. BMLs are recognisable on magnetic resonance imaging (MRI) data as excessive water signals while they are not visible in plain radiography (Figure 4) [38]. BMLs can be defined as common subchondral bone alternations in MR images among patients with knee OA. Multiple studies showed the association of BMLs and OA progression [36, 39, 37, 40, 41].

According to MRI Osteoarthritis Knee Score (MOAKS), each sub-region is separately assigned a number from 0 to 3 based on the percentage of areas of BMLs [42]. BMLs may appear in one or more sub-regions. The more detailed information is provided in Table 1.

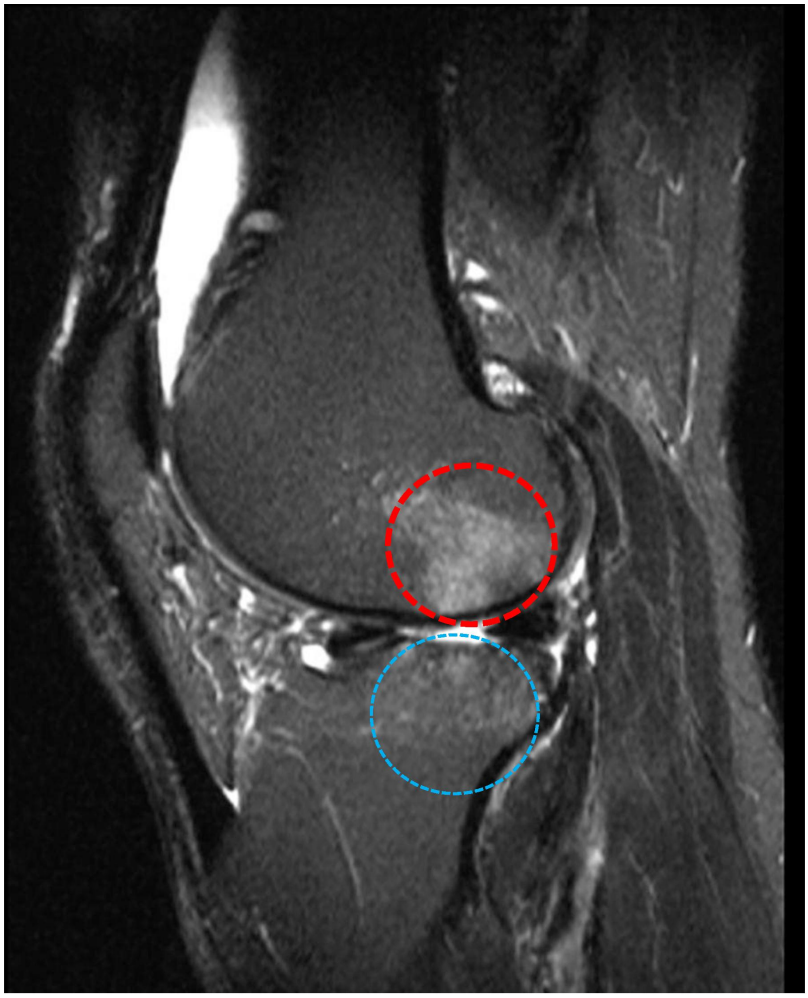


Figure 4. An illustration of intense bone marrow lesions (BML) at femur (red circle) and scattered BML in tibia (blue circle) in a MR image (Sagittal T2 fat-saturated).

Table 1. Scoring system for bone marrow lesions (BML)

Size of BML by volume	**BML score**
None	0
Less than 33% of subregional volume	1
Between 33-66% of subregional volume	2
More than 66% of subregional volume	3

1.4. Clinical Imaging

OA diagnosis primarily involves physical examination, review of medical history, and identification of symptoms [43]. Typical symptoms of OA include gradual onset of pain, morning stiffness with pain, restriction of movement, crepitus during joint

movement, joint effusions, joint deformities and subluxations [44]. Medical imaging modalities, particularly, plain radiography are commonly used to confirm the presence and quantify severity of OA [43]. Plain radiography of a knee joint is typically captured in a fixed-flexion standing position with fixed setting such as beam angle (Figure 7). Severity of OA can be assessed from plain radiographs by the Kellgren-Lawrence (KL) scoring system. There are many disadvantages of plain radiography, such as ionising radiation while performing radiography, providing 2-D projection image from 3-D tissue, insensitivity to early changes of OA, incapability in capturing soft tissues like articular cartilage and menisci, etc [45]. In spite of these drawbacks, plain radiography has been primarily used for OA imaging due to several advantages, such as being cost-efficient, fast, and broadly available [46].

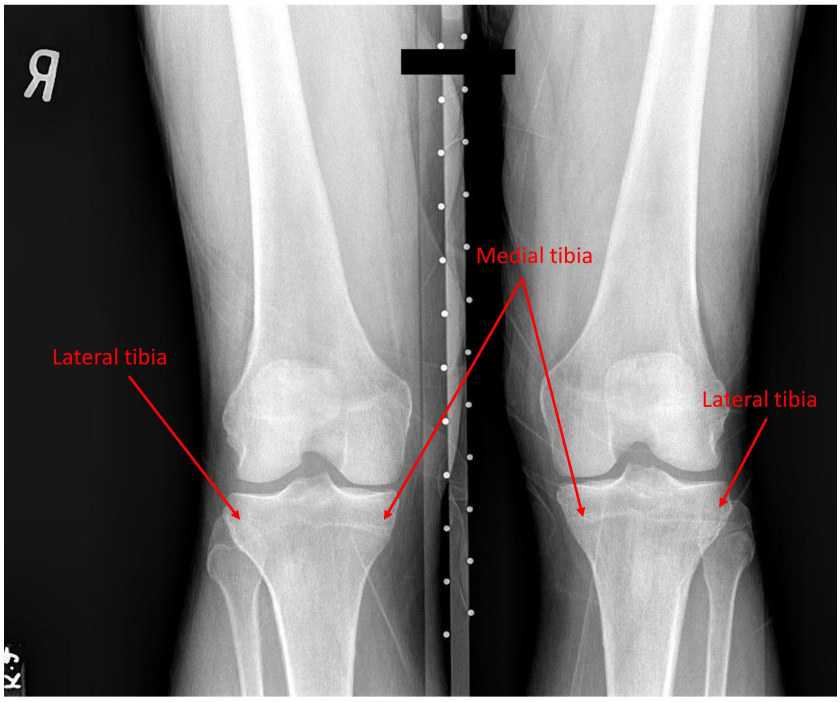
MRI is another clinically available modality that provide complementary and more detailed information such as morphology and composition of cartilage, ligaments, etc. As such, those modalities can be used for OA diagnosis and also OA research purposes [47]. MR images are produced by creating strong magnetic fields and radio waves. Compared to plain radiography, MRI does not have ionising radiation. MRI can provide multiple quantitative and semi quantitative features which potentially involved in OA such as BMLs, joint effusion, synovitis, etc. that are not visible in plain radiographs [48]. Because of these properties, MRI for knee joint in OA studies and also clinical practices is perceived as the most valuable imaging modality [49] (Figure 6).

1.5. Assessment of Osteoarthritis Severity

1.5.1. Kellgren-Lawrence Grading System

The first classification scheme for radiographic OA was presented in 1957 by Kellgren and Lawrence [50]. The proposed scheme was established for classifying OA from plain radiographs in eight joints including knee joint. The aim of such scheme was to provide a tool to differentiate the radiographic stages of the disease [51]. Having such classification scheme might greatly help healthcare institutions in terms of setting up treatment strategy for OA patients [52].

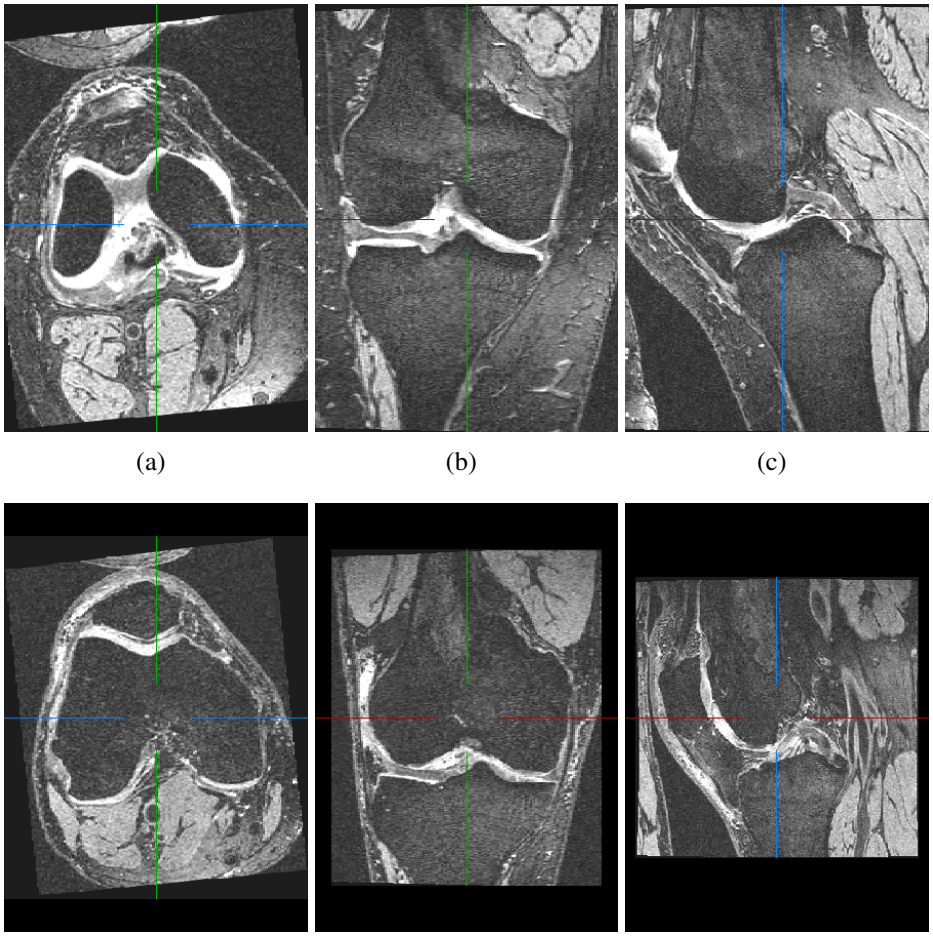
According to the KL grading system, a grade, ranging from 0 to 4, is assigned to a knee joint. Higher grades correspond to more severe stages of OA [50]. The grades represent: no signs of OA (KL-0), doubtful OA (KL-1), mild OA (KL-2), moderate OA (KL-3), severe OA (KL-4). The detailed explanation of each grade is provided in Table 2.



(a)

(b)

Figure 5. Two plain radiographs of knee joints of different subjects from the Osteoarthritis Initiative (OAI) dataset at the first visit. They are captured in a fixed-flexion standing position. Lateral and medial of tibia are illustrated in (a).



(a) (b) (c)

(d) (e) (f)

Figure 6. Two examples of magnetic resonance imaging (MRI) data, 3D dual echo steady state sequence (3D DESS) from OAI dataset. Figure shows three planes: axial plane (a,d), coronal plane (b,e) and sagittal plane (c,f).

Table 2. The description of Kellgren-Lawrence (KL) classification system

KL grade	**Description**
0	No radiographic features of osteoarthritis
1	Possible joint space narrowing and osteophyte formation
2	Definite osteophyte formation with possible joint space narrowing
3	Multiple osteophytes, definite joint space narrowing, sclerosis, and possible bony deformity
4	Large osteophytes, marked joint space narrowing, severe sclerosis, and definite bony deformity

1.5.2. MRI Osteoarthritis Knee Score (MOAKS)

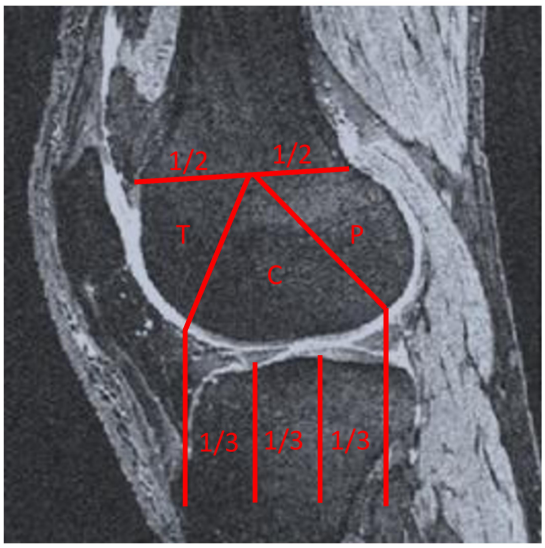
Whole-Organ MRI Score (WORMS) and Boston Leeds OA Knee Score (BLOKS) were developed as the first methods to assess knee OA status from MR images

[53][54]. Although both tools are broadly used in clinical practice and in OA research, there are several limitations in using these methods. For instance, WORMS meniscal scoring method combines several different factors and, in BLOKS, BML scoring is extremely time-consuming and complicated [42] . In order to address these limitations, MOAKS was proposed by Hunter et al. [42]. In MOAKS, knee joint is divided into 14 compartments including patella (medial patella and lateral patella), femur(medial and lateral trochlea, medial and lateral central femur, medial and lateral posterior femur), medial tibia (anterior, central and posterior), lateral tibia (anterior, central and posterior), and tibial spines. Then, different individual features from various tissues such as BML from trabecular bone, cartilage loss from AC, osteophyte from margins of joint, synovitis from synovial tissue, meniscal extrusion, meniscus tear, meniscal signal, and either absent or present anterior cruciate ligament tears, are then graded. These features are mainly defined because they are informative and reproducible enough in relation to OA progression and pain [42].

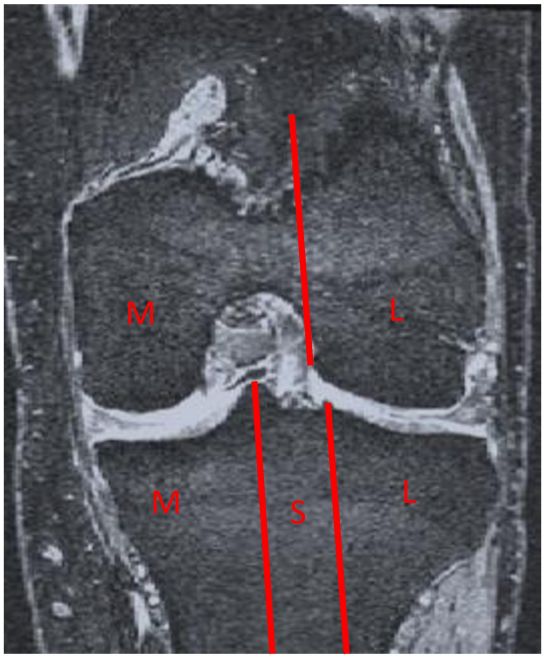
1.6. Related Studies

In OA research, there is a growing interest in subchondral bone alternation [55]. Currently, subchondral bone plays a critical role in diagnosis of musculoskeletal conditions such as OA [56]. Such changes in subchondral bone can be detected through textural analysis. Several studies have been performed in order to investigate whether OA incidence or progression is associated with textural changes in subchondral bone [57, 45, 58, 36, 39, 59, 60]. For instance, Janvier et al. [61], performed texture analysis from medial tibia using plain radiographs in order to predict knee OA progression. Their findings showed that trabecular bone textural parameters has a predictive ability to diagnose knee OA. A paper from Hirvasniemi et al. [60] investigated the connection between the differences in bone texture and radiography-based bone density among controls, subjects with knee OA and subjects with BMLs by using minimal and clinical post-processing. Their results showed that bone density and texture differences can be evaluated from plain radiographs when using minimal post-processing. Another study by Hirvasniemi et al. [57] was conducted in attempt to quantify the differences in medial and lateral sides of subchondral bone plate, trabecular bone, and femur by using plain radiography. Their results indicated that textural features were more robust and reproducible than bone density-related parameters when imaging conditions for all patient are different.

Apart from texture analysis from plain radiographs, several studies performed nearly the same statistical analysis on different MRI sequences. For example, MacKay et al.



(a)



(b)

Figure 7. All sub-regions defined by MRI Osteoarthritis Knee Score (MOAKS), indicated in sagittal (a) and coronal (b) views. In sagittal view (b), femur is divided into three region of trochlea (T), central (C), and posterior (P). Tibia has three equal sub-regions including anterior, central and posterior. In coronal view (b), tibia is divided into medial, subspinous (SS) and lateral sub-regions while femur is halved into the medial and lateral femoral condyle

[36] analysed whether MRI subchondral bone textural changes over 12–18 months can be used as an effective biomarker to predict OA progression over 36 months. Their results revealed that there is a significant association between initial and 12–18-month

MR textural changes in subchondral bone and OA progression at 36 months. Earlier study from MacKay et al. [39], showed that texture analysis of tibial subchondral bone from MR images has a potential role to play in OA pathogenesis and, thus, can be potentially used in treatment decision making. Their method yielded 97% of accuracy in the OA classification.

Some of the recent studies discovered that there are other causes apart from OA incidence which involved in textural changes in subchondral bone [62, 58]. For instance, Wolski et al. [58], compared trabecular bone texture among subjects with and without cartilage defects. Their results indicated that there are considerable changes in trabecular bone texture between these two groups. Similarly, a paper by Hirvasniemi et al. [62] which is the closest work to our study, demonstrated that subchondral bone structure in plain radiography can have significant differences in subjects with and without AC damage or BMLs. In their method, textural analysis were performed on 80 subjects with various OA severity. The results showed that there is association between subchondral bone structure among subjects with and without cartilage damage or BMLs. But in our study, besides the aforementioned analysis, we also investigate the relationship between BMLs and presence of OA.

As mentioned previously, textural analysis of subchondral bone has quite critical role in OA studies but it is also crucial to investigate which locations of subchondral bone contain valuable textural information. Bayramoglu et al. [59] investigated the impact of region of interest (ROI) location in subchondral bone in plain radiography on texture analysis. They employed several texture descriptors to classify subjects with and without OA. They presented a fully automatic algorithm to extract the most influential region of subchondral bone by utilising their proposed adaptive segmentation. They segmented bone region into several sub-regions and performed texture analysis independently on each sub-regions. As a result, they showed that selecting an adaptive ROI had positive effect on binary classification performance in comparison with a standard ROI in subchondral bone used in prior studies in OA detection. In fact, they reached to the conclusion that marginal sub-regions of subchondral bone in tibia and femur have significant textural changes between subjects with OA and without OA compared to other regions.

As the previous studies confirmed that textural changes of subchondral bone are associated with OA progression and also with BMLs, it can be persuasive enough to conduct more works to investigate BMLs contribution with OA incidence by analysing subchondral bone texture.

The main focus of this thesis is to explore the effect of BML in OA diagnosis from the subchondral bone texture by using plain radiography. To do so, we divided subjects into three groups in accordance with their differences in BML presence and BML

location. Then, statistical comparisons were made to discover whether BMLs can make significant differences between OA and non OA subjects. Additionally, more comparison were performed in order to discover the relationship between the presence of BMLs and the location of textural changes affected by BMLs.

As a subsidiary aim of this thesis, we desired to implement deep convolutional neural network (CNN) for OA detection by using subchondral bone texture from plain radiography and MRI. The reasons behind this aim is to firstly investigate whether our CNN model is able to detect OA when subjects have no BMLs. The second second is to compare the subchondral bone textural features of plain radiography and MRI in OA detection.

2. METHODS AND MATERIALS

2.1. The Osteoarthritis Initiative (OAI) Dataset

This study uses data from The Osteoarthritis Initiative (OAI) (publicly available at https://nda.nih.gov/oai). OAI is a prospective longitudinal study focusing primarily on knee OA. It includes the medical images from seven imaging follow-up visits (baseline, 12 months, 24 months, 36 months, 48 months, 72 months, 96 months) from three subgroups: 1) individuals who are at risk of OA progression, 2) individuals who are at high risk of OA progression, and 3) a normal control group. The total of 4796 individuals with age range of 45 to 79 participated in the study. In this thesis, the baseline visit was merely used. At the baseline, all subjects underwent weight-bearing posteroanterior fixed-flexion knee radiographic scanning (with beam angle of 10°). The data acquisition was performed across the four centres. Nearly all the plain radiographs (n=1974) at the baseline were graded based on semi-quantitative KL scoring systems, while MOAKS BML scoring from MR images at the baseline was not available for all subjects. The MOAKS scoring used in this thesis is provided by Boston Imaging Core Lab [63]. Knee MRI section of OAI at the baseline contains the following sequences:

- Sagittal 3D dual echo steady state with Water Excitation (SAG 3D DESS WE)
- Coronal Intermediate-Weighted 2-D Turbo Spin-Echo (Coronal IW 2D TSE)
- Sagittal Intermediate-Weighted 2D Turbo Spin-Echo Fat Suppressed (Sagittal IW 2D TSE Fat Suppressed)
- Sagittal 2D multi-echo spin-echo (Sagittal 2D multi-echo SE)

2.2. Study Design

There are two sub-studies presented in this thesis:

1. The first and main sub-study started with selecting subjects from OAI dataset in accordance with BMLs scores. Three groups based on the BML score and location were determined. Subsequently, preprocessing techniques, including ROI extraction, were employed. After extracting the ROIs, several textural features in were computed from the ROI patches. Ultimately, statistical analysis for establishing interpretable comparison between these three groups were applied. The organisation of our method is elaborated in the following flowchart (Figure 8).

2. In the second sub-study, as a proof of concept, we utilised a CNN model to classify subjects with and without OA by using the subchondral bone textures from their 3-D DESS MRI and plain radiography. We only selected subject with no BMLs in order to eliminate the effect of BMLs on subchondral bone structure. In fact, our aim was to investigate whether our CNN model is able to distinguish textural differences at subchondral bone between OA and non OA subjects. Furthermore, we wanted to discover which modality has more capability to capture textural differences among these two classes. Thus, we conducted the experiments separately for MR images and radiographs.

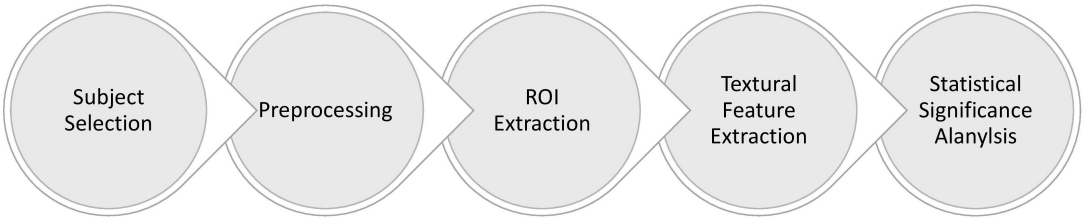


Figure 8. The pipeline of the first sub-study contains 5 steps. The pipeline starts with subjects selection, then implementing preprocessing (knee joint localisation, image normalisation, image alignment), ROI extraction, feature extraction by 4 sets of texture descriptors and finally culminate with statistical analysis.

The methods and results of the first sub-study is presented in the main text. The second sub-study is organised in the Appendix as follows: subject selection, preprocessing for MR images, networks and implementation, and finally results. The summary of both sub-studies are shown in Table 3.

Table 3. Specification of the sub-studies implemented in the thesis. CNN - convolutional neural network, FD - fractal dimension, BML - bone marrow lesion, GLCM - Gray-Level Co-Occurrence Matrix

Sub-study	**Material**	**Feature descriptors**	**Method**
First	Subjects with and without BMLs	GLCM, absolute gradient, histogram, FD	Statistical significance tests (Fisher and Mann Whitney)
Second	Subjects without BMLs	CNN	Binary classification

2.3. Study Subjects

We selected subjects from the baseline of the OAI (see Figure 9). Then, subjects with missing assessment information such as BML scores and KL grades were excluded. The selected subjects were categorised into three subject groups. The group 1 consisted of subjects who had non-zero BML (BML score of 1,2 or 3) at least in medial tibia sub-region, including posterior, anterior, or central. Other sub-regions in this group were ignored and they might have or not have BMLs. The group 2 included subjects without BMLs (BML score of 0) in femur and tibia in all sub-regions. Finally, the group 3 had the subjects who did not have BMLs in medial tibia, but had BMLs, at least, in one of the other sub-regions (see Table 4 and Figure 13).

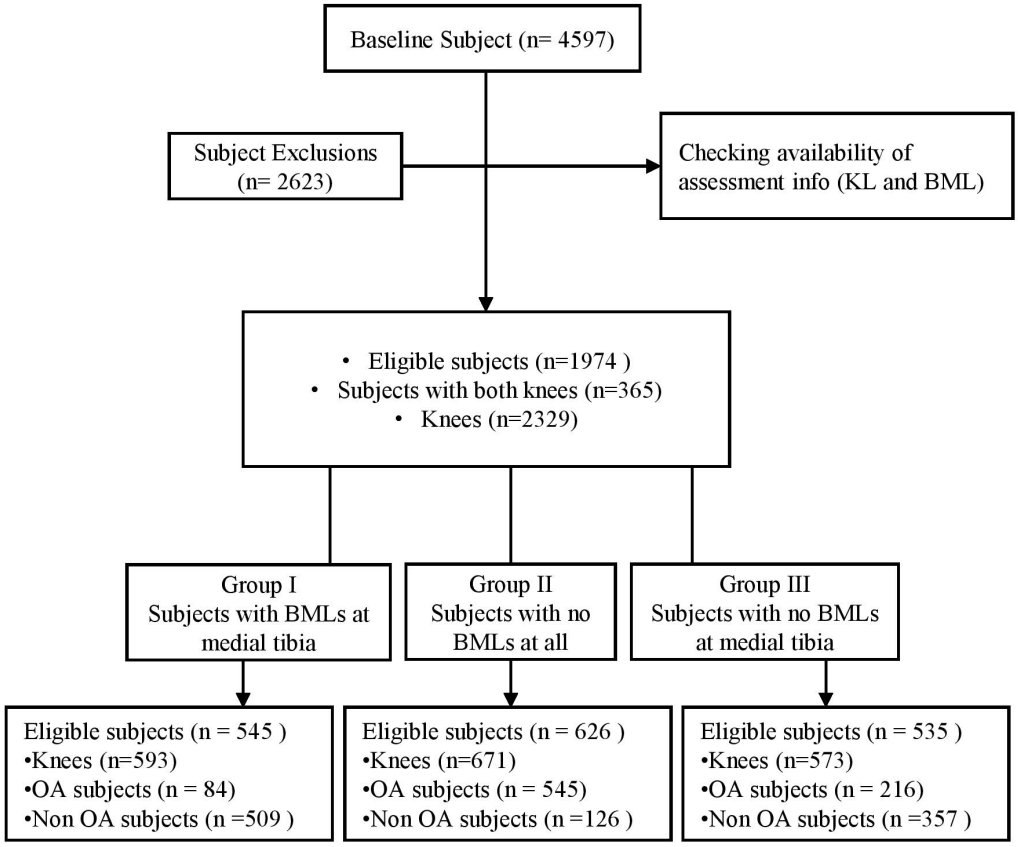


Figure 9. The flow-diagram for selection of study participants and dividing subjects into three groups based on bone marrow lesion (BML) scores. OA - osteoarthritis, KL - Kellgren and Lawrence system

Table 4. Clarification of the considered subject groups in terms of BML presence

BML locations	**Group 1**	**Group 2**	**Group 3**
Tibial media	BML	No BML	No BML
Other sub-regions	Might have	No BML	Might have

2.4. Data Preprocessing: Plain Radiographs

The following preprocessing pipeline utilised on the plain radiographs composed of image normalisation, knee joint localisation, and image alignment. Afterwards, the ROI in subchondral bone was selected and then extracted. The codes of this pipeline were provided by the Research Unit of Medical Imaging, Physics, and Technology (MIPT), Oulu, Finland (https://github.com/MIPT-Oulu).

1. **Normalisation:** The histograms of 16-bit DICOM images (plain radiographs) were truncated between the 5th and 99th percentiles and a global contrast normalisation is applied to the images. Finally the images are converted to 8-bit images.
2. **Knee Joint Localisation:** Knee joint localisation is performed based on the anatomical landmarks provided by BoneFinder as it is done in [59, 22]. BoneFinder is a machine learning-based and fully automatic software tool to outline and segment skeletal structure from 2D radiographs designed by the Centre for Imaging Sciences at The University of Manchester, Manchester, UK [64]. Firstly, the centre of the joint is determined by using the marginal landmarks from tibia. Then, the width of the knee patches were calculated depending on the spacing of each image. In this work, the width of knee patches is 140 mm. Finally, the knee joints of both the left and the right legs were extracted assuming that centre of the joint is the centre of the extracted area.
3. **Image Alignment:** Knee joints are not aligned from a subject to another. Misaligned knee joint can effect negatively on interpretation of textural features extracted from subchondral bone. Therefore, each knee joint image should be rotated so that the line connecting the margins of knee joint becomes horizontal. Rotation angle is calculated accordingly. All landmarks are then transformed appropriately.
4. **ROI Extraction:** There has been several ROI extraction methods implemented in this study but we eventually chose marginal region of medial tibia as previous study suggested [59]. To extract marginal ROI, fixed size of 64x64 square pixels was examined in which its top right vertex is overlapped with marginal landmark

in medial tibia (see Figure 11). Before preprocessing, left knee images were flipped horizontally.

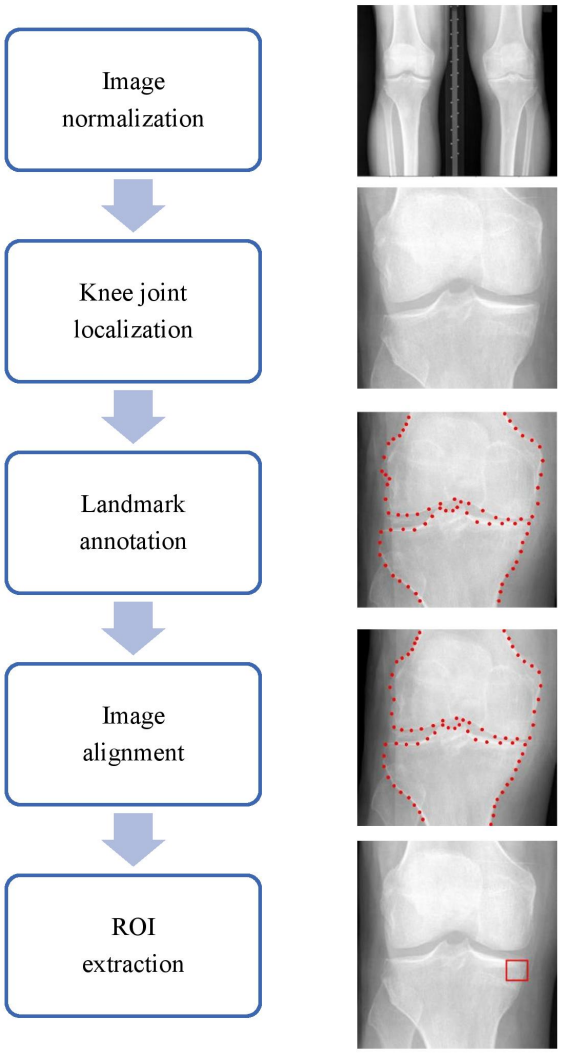


Figure 10. Preprocessing pipeline for plain radiographs. Histogram truncated and global contrast normalisation was applied to images in the first stage. Then knee joint was located and extracted. Afterwards, landmarks were annotated by BoneFinder software. By using marginal landmarks, all knee joint images were rotated to be aligned. Finally, region of interest (ROI) from margins of medial tibia were extracted.

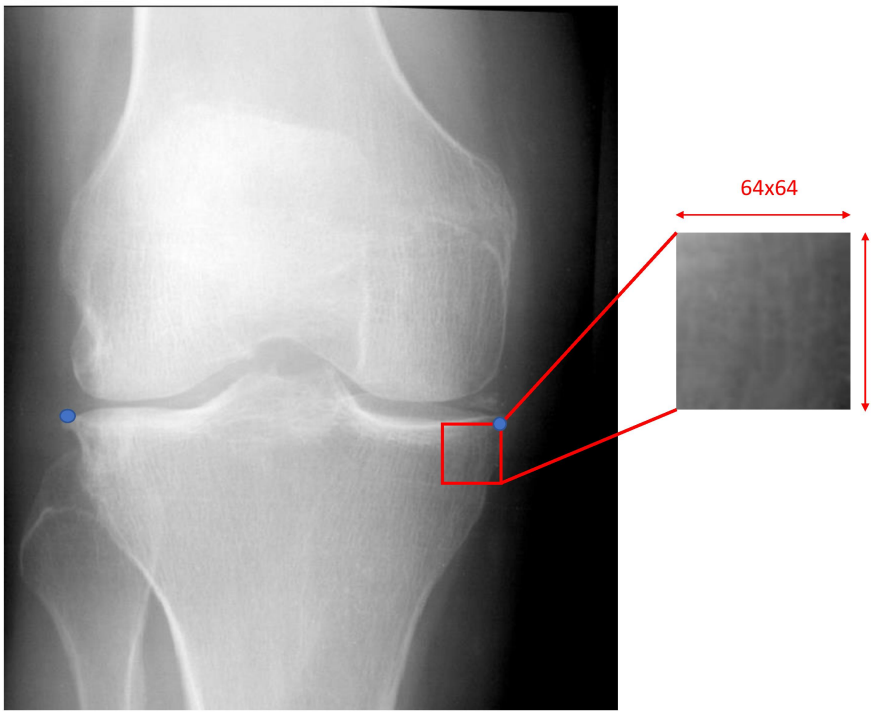


Figure 11. The figure presents the extraction of marginal region of interest (ROI) from medial tibia in plain radiograghs

2.5. Texture Analysis

Features of subchondral bone textures at medial tibia were extracted by four textural descriptors including Grey-level co-occurrence matrix (GLCM), Image histogram , Absolute gradient and fractal dimension (FD). In this section, we are going to explain these descriptors and their features that we used in the statistical analysis.

1. **Grey-level co-occurrence matrix (GLCM):** GLCM is one of the first methods proposed for extraction of textures. It was proposed for the first time by Haralick back in 1973 [65]. Since then, it has been commonly used for wide range of applications and studies that are in the domain of texture analysis. Several texture features based on GLCM were presented in various literature [66, 67, 68, 69].

   Generally, GLCM is effective when the size of the texture samples is small because of high dimensionality of the matrix and high processing time and complexity [70]. Furthermore, it has been proved that the GLCM features from images with a large amount of noises are not robust [70]. Due to the high resolution of plain radiographs and small size of textures, we perceived GLCM as one of the suitable textural descriptor in this study.

GLCM can capture grey level spatial dependencies in different orientations and distances by establishing a co-occurrence matrices. Based on the chosen orientation and distances and GLCM can extract useful textural information from its features [71]. The GLCM matrix is defined as a square matrix (G) of size of $N \times N$ where N is the total number of intensity in the image. The element of $g^{i,j}$ in matrix G represents the number of times a pixel A with intensity value i is separated from a pixel B with intensity value j at a particular distance r in a particular direction d. GLCM parameters used in this thesis are as follows: distances = 1 to 3 pixles, angles = 0, 45, 90, 135 degrees. Since we have 3x4 parameters, GLCM is calculated for 12 times. Finally, the mean value of all 12 computed GLCM was used for analysis.

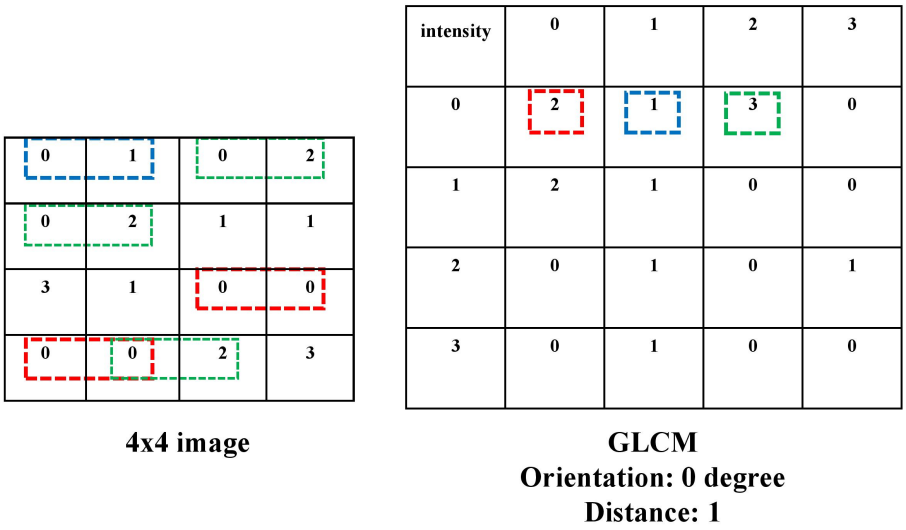


Figure 12. An illustration of GLCM with 4x4 image having intensity range of 0 to 3. here, the orientation is 0 degree and the distance is 1.

We used six GLCM texture parameters including dissimilarity, contrast, correlation, Entropy, Homogeneity, Uniformity or angular second moment (ASM).

(a) **Contrast:** This parameter indicates the amount of variation in local gray-level values. Higher contrast values indicate presence of edges, noise, or wrinkled textures in the image.

(b) **Homogeneity:** It represents the smoothness of the the gray level distribution of the image. It is roughly opposite of contrast. For instance, if the contrast is small, usually homogeneity is large or vice versa.

(c) **Correlation:** It measures the linear dependency of gray intensities on those of neighbouring pixels; it provides a measure similar to autocorrelation methods.

(d) **Uniformity (ASM):** It estimates the uniformity or orderliness of the gray-level distribution in the image. Images with a smaller number of gray levels have larger uniformity.

(e) **Entropy:** It measures the degree of disorder among pixels in the image; it is inversely correlated with uniformity; images with a larger number of gray levels have larger entropy.

2. **Absolute Gradient:** It is a very common operator for edge detection is the Sobel operator. The Sobel is a discrete differentiation operator which computes the convolution of an image with kernel D of odd sizes. 3x3 kernel was chosen in this thesis. Horizontal and vertical structural changes can be obtained as follows:

$$D_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

$$D_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

$$G_x = I * D_x \tag{1}$$

$$G_y = I * D_y \tag{2}$$

G_x and G_y are vertical and horizontal derivative approximations, respectively. I is the input image and D_x and D_y are horizontal and vertical kernels.

Another edge detector named Laplacian was also used in this work. Unlike the Sobel operator, Laplacian calculates the second derivatives and uses a single kernel. Before applying Laplacian filter, Gaussian filter is usually applied to smoothen the image in order to eliminate the noises. Laplacian operator for 2-D grey-scale images is defined as follows:

$$L(x, y) = \frac{\varphi^2 I}{\varphi x^2} + \frac{\varphi^2 I}{\varphi y^2} \tag{3}$$

Features such as mean, variance, skewness, kurtosis and number of pixels with non-zero gradient (NPWNZG) were computed from resulted images.

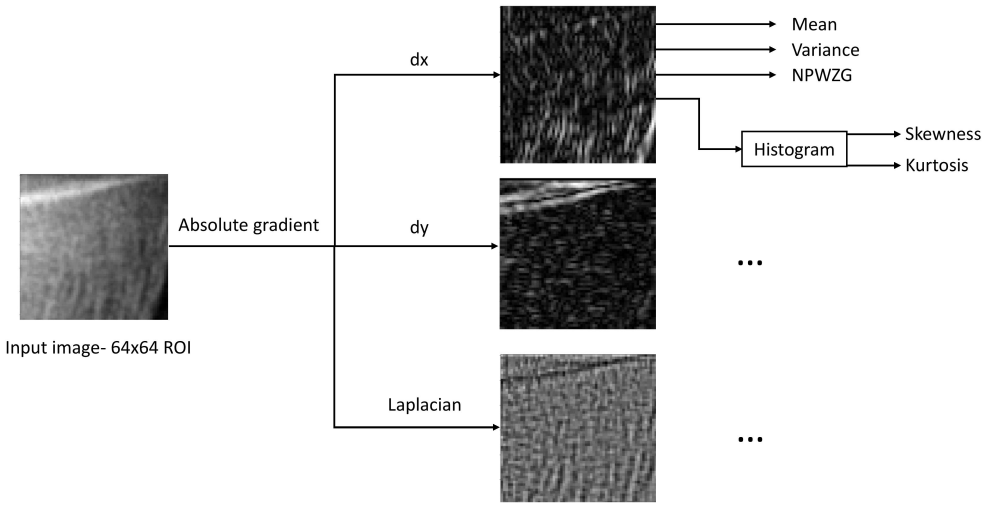


Figure 13. An illustration of extracting absolute gradient features. Firstly, the gradient (horizontal - dx, vertical - dy, and laplacian) is computed from the input image. Then features including mean, variance, and number of pixels with non-zero gradient (NPWNZG) were obtained from the output image. Other features including skewness and kurtosis were computed from the histogram of the output image. These features were separately estimated from the dx, dy, and laplacian.

3. **Fractal Signature Analysis (FSA):**

   Fractal signature analysis (FSA) [62, 72, 73, 74] is an approach for measuring fractal dimension (FD) which is connected with the roughness and complexity of the image. High values of FD represent the high complexity of the image. FD has been frequently used for textural analysis particularly for high resolution images such as radiographs [62, 72, 73, 74, 75].

   In FSA, fractal dimension of an image is estimated separately in vertical and horizontal directions over various range of scales. In this thesis, the image firstly underwent morphological transformations including dilation and erosion. Subsequently, a base-ten logarithm of the differences between dilated and eroded of the image was computed. This calculation was repeated with different values of length r from 2 to 16 pixels.

   The surface A(r) is calculated as follows:

$$A(r) = \frac{V(r) - V(r-1)}{2} \tag{4}$$

Fractal dimension is obtained from the slope of the line $A \cdot log(r)$. Eventually, 11 FD features was generated. Each of them represents FD in one scale. We chose the first and last scale of FD for texture analysis (in the following, defined as FD ver/hor 1 and FD ver/hor 2, respectively). The slope of two consequent values of r is is defined by:

$$S(r) = \frac{Log(A(r)) - Log(A(r-1))}{Log(r) - Log(r-1)} \tag{5}$$

4. **Histogram**: Intensity histogram shows the distribution of gray-level values of an image. In this thesis, four features including skewness, mean, variance, and kurtosis were computed from the intensity histogram of the images.

2.6. Statistical Analysis

Two statistical analysis were used to investigate how the textural features change between OA and non OA subjects when subjects have or not have BMLs (first analysis) and how those changes differ across the groups (second analysis). In the first analysis, all aforementioned textural features from extracted ROIs were computed. Then, each feature was compared between the OA and non OA subjects in each group by using two statistical significant tests including Fisher's exact test and Mann–Whitney U test. In fact, these tests were used in order to investigate which textural features are significantly different between two classes of OA and non- OA subjects in each group (OA vs. non OA in group 1, OA vs. non OA in group 2, OA vs. non OA in group 3).

The aim of using Fisher's exact test here, is to examine whether there is a significant association between the variables of two classes in a contingency tables such as confusion matrix. Moreover, Mann–Whitney U test was used to discover how data of two classes (OA vs. non OA) were differently distributed. Generally, T-test is used for investigating significant differences between two independent data if the data have normal distributions. Instead, we used Mann–Whitney U test because our data was not normally distributed. Normality of each texture feature was evaluated by Shapiro-Wilk Test. For Mann–Whitney U test and Fisher's exact, p-value under 0.05 was considered to be statistically significant.

For Fisher's exact test, each textural feature was individually assessed by employing binary classification (OA vs. non-OA). In binary classification, our model was trained by the all data and all data were tested afterwards. Then, p-value and diagnostic odds ratio (OR) were estimated from the resulted confusion matrix obtained from

classification. Eventually, each feature has two values of p-value and OR. p-value under 0.05 was perceived as significant and OR greater than 1 shows how strong the association between two classes is. By looking at these values, significant features can be identified among OA and non OA classes (OA = KL grade of 0, 1 and non OA = KL grade of 2, 3, and 4).

For Mann–Whitney U test, the distribution of each textural feature for subjects with and without OA was individually checked. Finally, a p-value from each textural feature is obtained. The p-values determined whether the feature is significantly different between OA and non-OA subjects. We should note that these tests were implemented separately for all subject groups including subjects with BMLs in medial tibia , subjects without BMLs and subjects with no BMLs in medial tibia.

In the second analysis, we performed Mann–Whitney U test once more between the subject groups. In this test, textural features of one subject group was compared to the textural features of another subject group. For instance, features of all subjects with BMLs at least in medial tibia (group 1) was compared to features of all subjects without BMLs (group 2). Such comparison was made between group 1 vs. group 2, group 1 vs. group 3 , and group 2 vs. group 3. Finally, results from this tests examined which features between different subject groups were statistically different. The interpretation of the results from these tests can determine whether presence of BMLs can locally or globally effect on OA diagnosis.

3. RESULTS

In this section, results from statistical analysis are provided. In the first part of analysis, results from comparing textural differences among OA and non OA in each subject group separately were obtained to uncover the effect of BML presence in OA detection (section 3.1). In the second part of analysis, comparisons between the three subject groups were performed to investigate the impact BML location in subchondral bone texture (section 3.2). In this analysis, 29 textural features from 4 descriptors were assessed by both Fisher's Exact Test and Mann-Whitney U Test. After employing these tests, significant differences between features of subjects with and without OA were identified in each group (BMLs at medial tibia, no-BML and no-BML at medial tibia). In the second part of analysis, same features were compared across the groups by Mann-Whitney U Test.

3.1. Group-Wise: OA Vs. Non OA

In Table 5, all textural features for subjects with BMLs in medial tibia are shown. As it is illustrated, 19 features from the total 29 features were significantly different by acquiring p-values less than 0.05 from Fisherś exact test. Results from Mann-Whitney U test also showed that these 19 features between OA and non-OA were differently distributed. Additionally, features such as mean-dx ($p=0.059$), contrast ($p=0.077$), skewness-laplacian ($p=0.067$), variance-laplacian ($p=0.077$) showed weakly significant differences between OA and non OA. They also have different distributions according to Mann-Whitney U test (see Table 5).

Results of Fisherś exact test for subjects without BMLs, however, was different from subjects with BMLs (see Table 6). Only 5 features including variance-dy ($p=0.037$), NPWNZG-dx ($p=0.049$), histogram skewness ($p=0.008$), and histogram kurtosis ($p=0.022$) had significant differences between OA and non OA subjects. All of them have different distributions according to Mann-Whitney U test. Among these features, only histogram skewness obtained p-value less that 0.01.

Within subjects without BMLs in medial tibia (group 3), there were almost no significant differences between subjects with and without OA (Table 7). Although features such as mean-laplacian, FD vertical 1 and 2 and FD horizontal 1 have different distributions between subjects with and without OA according to the Table 7, they had no significant differences among OA and non OA according to Fisher test.

Table 5. Comparison of subchondral bone textural features between subjects with and without osteoarthritis (OA). Considered subjects had bone marrow lesions (BML) in medial tibia (group 1). Values are: columns 2 and 3 - mean (standard deviation), columns 4 and 5 - p value of Fisher's Exact Test and Mann-Whitney U Test, respectively. NPWNZG - number of pixels withnon-zero gradient, GLCM - gray-level co-ocurrence matrix, ASM - angular second momentum, FD - fractal dimension.

Feature	**No OA (SD)**	**OA (SD)**	p **value** Fisher test	p **value** U-test
Absolute Gradient				
Mean (dx)	21.755 (6.407)	19.763 (7.469)	0.059	**0.009**
Mean (dy)	18.690 (5.773)	16.787 (6.449)	**0.034**	**0.004**
Mean (laplacian)	0.315 (0.360)	0.243 (0.326)	**0.006**	**0.021**
Variance (dx)	378.357 (234.411)	363.924 (259.171)	0.721	0.171
Variance (dy)	398.828 (267.226)	285.792 (233.872)	**0.002**	**<0.001**
Variance (laplacian)	1282.119 (664.836)	1087.968 (754.599)	0.077	**0.002**
Skewness (dx)	2.479 (0.509)	2.750 (0.702)	**0.006**	**<0.001**
Skewness (dy)	2.897 (0.632)	3.003 (0.723)	0.109	0.193
Skewness (laplacian)	3.317 (0.910)	3.615 (1.117)	0.067	**0.014**
Kurtosis (dx)	6.222 (5.045)	7.661 (4.475)	**0.013**	**<0.001**
Kurtosis (dy)	8.376 (4.932)	8.969 (4.989)	0.1	0.170
Kurtosis (laplacian)	11.243 (8.041)	13.757 (9.734)	**0.035**	**0.017**
NPWNZG (dx)	0.952 (0.007)	0.948 (0.011)	**0.011**	**0.001**
NPWNZG (dy)	0.947 (0.010)	0.944 (0.012)	0.087	0.079
NPWNZG (laplacian)	0.969 (0.016)	0.963 (0.021)	**0.035**	**0.011**
GLCM				
Dissimilarity	5.709 (1.660)	5.161 (1.942)	**0.033**	**0.007**
Correlation	0.934 (0.032)	0.939 (0.039)	**0.009**	0.126
Contrast	64.433 (33.868)	55.212 (36.585)	0.077	**0.006**
Homogeneity	0.188 (0.060)	0.214 (0.085)	**0.037**	**0.011**
ASM	0.001 (<0.001)	0.001 (0.001)	**0.012**	**<0.001**
Energy	0.030 (0.006)	0.033 (0.010)	**0.006**	**<0.001**
Histogram				
Mean	162.181 (25.766)	171.263 (32.240)	**0.019**	**<0.001**
Variance	565.147 (241.371)	545.445 (363.842)	0.333	**0.02**
Skewness	2.006 (3.548)	2.606 (2.318)	**0.005**	**<0.001**
Kurtosis	1.643 (0.543)	1.837 (0.547)	**0.003**	**<0.001**
Fractal Dimension				
FD_ver_scale 1	2.805 (0.162)	2.790 (0.167)	**0.044**	0.07
FD_ver_scale 2	2.635 (0.293)	2.643 (0.290)	**0.003**	0.29
FD_hor_scale 1	2.736 (0.124)	2.735 (0.138)	0.814	**<0.001**
FD_hor_scale 2	2.781 (0.240)	2.655 (0.245)	**<0.001**	**0.36**

3.2. Differences Between Subject Groups

Results of Mann-Whitney U test employed as the second part of statistical analysis, show that subjects with BMLs at medial tibia have significant differences compared to subjects with no-BML and subjects with no-BML at medial tibia (group 1 vs. group2

Table 6. Comparison of subchondral bone textural features between subjects with and without osteoarthritis (OA). Considered subjects did not have bone marrow lesions (BML) (group 2). Values are: columns 2 and 3 - mean (standard deviation), columns 4 and 5 - p value of Fisher Exact Test and Mann-Whitney U Test, respectively. NPWNZG - number of pixels withnon-zero gradient, GLCM - gray-level co-ocurrence matrix, ASM - angular second momentum, FD - fractal dimension.

Feature	**No OA (SD)**	**OA (SD)**	p **value** Fisher test	p **value** U-test
Absolute Gradient				
Mean (dx)	21.249 (6.826)	20.439 (7.147)	0.552	0.130
Mean (dy)	19.025 (6.476)	18.327 (6.481)	0.424	0.143
Mean (laplacian)	0.358 (0.383)	0.310 (0.336)	0.552	0.164
Variance (dx)	356.374 (230.006)	360.026 (248.889)	0.553	0.426
Variance (dy)	436.981 (305.398)	377.114 (256.499)	**0.037**	**0.040**
Variance (laplacian)	1307.654 (928.381)	1159.486 (709.565)	0.766	**0.042**
Skewness (dx)	2.482 (0.441)	2.630 (0.654)	0.096	**0.011**
Skewness (dy)	2.917 (0.631)	2.990 (0.796)	0.532	0.378
Skewness (laplacian)	3.356 (0.960)	3.531 (1.159)	0.168	0.084
Kurtosis (dx)	5.948 (2.217)	7.118 (6.703)	0.091	**0.003**
Kurtosis (dy)	8.427 (4.113)	9.251 (7.856)	0.519	0.351
Kurtosis (laplacian)	11.472 (7.930)	13.306 (12.544)	0.195	0.100
NPWNZG (dx)	0.952 (0.007)	0.950 (0.011)	**0.049**	**0.021**
NPWNZG (dy)	0.947 (0.010)	0.945 (0.014)	0.526	0.258
NPWNZG (laplacian)	0.968 (0.017)	0.965 (0.020)	0.333	0.150
GLCM				
Dissimilarity	5.672 (1.860)	5.441 (1.877)	0.920	0.125
Correlation	0.943 (0.036)	0.944 (0.033)	0.553	0.472
Contrast	65.166 (39.418)	60.483 (35.927)	0.844	0.139
Homogeneity	0.192 (0.065)	0.202 (0.073)	0.170	0.110
ASM	0.001 (0.001)	0.001 (0.001)	0.261	**0.013**
Energy	0.029 (0.006)	0.031 (0.011)	0.268	**0.012**
Histogram				
Mean	157.163 (32.089)	155.844 (33.439)	0.839	0.32
Variance	663.593 (338.154)	619.970 (296.530)	0.316	0.09
Skewness	1.262 (2.785)	3.087 (15.028)	**0.008**	**<0.001**
Kurtosis	1.470 (0.505)	1.713 (1.026)	**0.022**	**<0.001**
Fractal Dimension				
FD_ver_scale 1	2.808 (0.178)	2.787 (0.163)	0.165	**0.02**
FD_ver_scale 2	2.583 (0.304)	2.611 (0.283)	0.116	0.49
FD_hor_scale 1	2.718 (0.119)	2.715 (0.119)	0.921	0.49
FD_hor_scale 2	2.750 (0.261)	2.711 (0.246)	0.322	**<0.001**

and group 1 vs. group 3). Subjects with BMLs at medial tibia and with no-BML (group 1 vs. group 2) have significant differences in 18 textural features. Likewise, 15 textural features were significantly different between subjects with BMLs at medial tibia and subjects with no-BML at medial tibia (group 1 vs. group 3). On the other hand, among subjects with no-BML and with no-BML at medial tibia (group 2 vs. group 3), there were almost no significant differences (see Table 9).

Table 7. Comparison of subchondral bone textural features between subjects with and without osteoarthritis (OA). All considered subjects have no bone marrow lesions (BML) in medial tibia (group 3). Values are: columns 2 and 3 - mean (standard deviation), columns 4 and 5 - p value of Fisherś Exact Test and Mann-Whitney U Test, respectively. NPWNZG - number of pixels withnon-zero gradient, GLCM - gray-level co-ocurrence matrix, ASM - angular second momentum, FD - fractal dimension

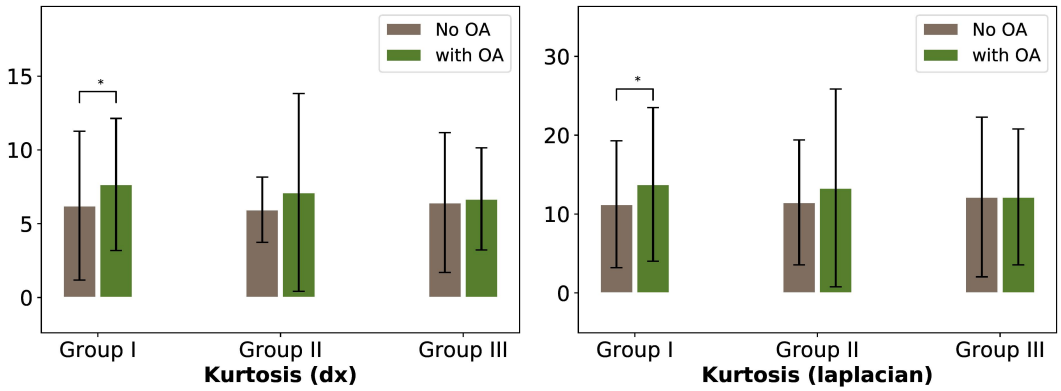
Feature	**No OA (SD)**	**OA (SD)**	p **value Fisher test**	p **value U-test**
Absolute Gradient				
Mean (dx)	20.641 (7.147)	20.665 (7.089)	0.931	0.393
Mean (dy)	18.564 (6.554)	19.212 (6.768)	0.605	0.113
Mean (laplacian)	0.314 (0.383)	0.365 (0.363)	0.069	**0.049**
Variance (dx)	342.595 (259.576)	369.441 (250.877)	0.083	0.076
Variance (dy)	427.671 (311.507)	451.898 (324.499)	0.728	0.204
Variance (laplacian)	1260.873 (889.004)	1213.507 (757.679)	0.730	0.477
Skewness (dx)	2.548 (0.543)	2.598 (0.550)	0.659	0.150
Skewness (dy)	2.997 (0.714)	2.952 (0.664)	0.719	0.188
Skewness (laplacian)	3.425 (1.024)	3.437 (0.974)	1.000	0.452
Kurtosis (dx)	6.435 (4.738)	6.680 (3.455)	0.464	0.138
Kurtosis (dy)	9.105 (6.351)	8.790 (4.739)	0.854	0.267
Kurtosis (laplacian)	12.163 (10.122)	12.161 (8.615)	1.000	0.444
NPWNZG(dx)	0.951 (0.009)	0.950 (0.009)	0.536	0.309
NPWNZG (dy)	0.945 (0.012)	0.946 (0.011)	0.719	0.229
NPWNZG (laplacian)	0.967 (0.018)	0.966 (0.017)	0.716	0.440
GLCM				
Dissimilarity	5.530 (1.910)	5.593 (1.888)	0.795	0.249
Correlation	0.944 (0.032)	0.937 (0.037)	0.342	**0.019**
Contrast	62.968 (40.987)	65.225 (39.887)	0.491	0.177
Homogeneity	0.198 (0.067)	0.198 (0.068)	0.589	0.409
ASM	0.001 (0.001)	0.001 (0.001)	0.319	0.060
Energy	0.030 (0.009)	0.031 (0.008)	0.082	0.059
Histogram				
Mean	153.652 (33.353)	154.764 (33.553)	0.598	0.32
Variance	640.251 (328.289)	601.498 (323.116)	0.428	0.09
Skewness	2.518 (11.778)	2.549 (5.078)	**<0.001**	**<0.001**
Kurtosis	1.622 (0.875)	1.754 (0.633)	**<0.001**	**<0.001**
Fractal Dimension				
FD_ver_scale 1	2.808 (0.178)	2.787 (0.163)	0.165	0.03
FD_ver_scale 2	2.583 (0.304)	2.611 (0.283)	0.116	**<0.001**
FD_ver_scale 1	2.808 (0.178)	2.787 (0.163)	0.165	**0.01**
FD_ver_scale 2	2.583 (0.304)	2.611 (0.283)	0.116	0.1

From another perspective, subjects across the groups can be compared in terms of the number of features which have significant differences between OA and non OA subjects. To elaborate more, 19 textural features from subjects with BMLs in medial tibia (group 1) were significantly different between OA and non OA while this number for subjects without BMLs (group2) and subjects without BMLs in medial

tibia (group3), were 4 and 2, respectively. Similarly, subjects with medial tibia BMLs (group 1) have relatively higher OR values than the other groups. In nearly all features, OR values for subjects with BMLs were higher that 1.5 while for subjects with no BMLs (group 2) and subjects with no BMLs at medial tibia (group 3) were almost below 1.5. To clarify more, the mean and standard deviation of each feature from both OA and No OA subjects in all groups are shown via error bar charts. Significant differences between OA and no OA subjects are specified with star (see Figure 14 and Figure 15).

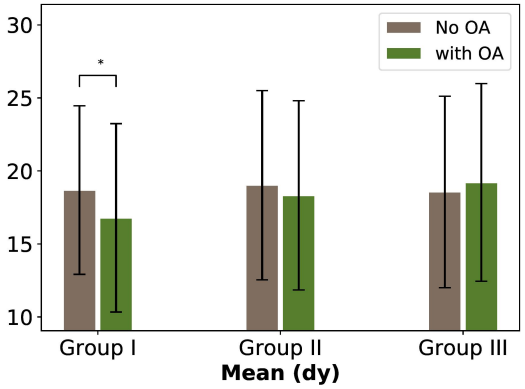
Table 8. Odds Ratio (OR) for each subject group was computed from binary classification based (OA vs. non OA defined by Kellgren and Lawrence (KL) system). Columns 2: OR of subjects with BMLs in medial tibia (group 1), columns 3: subjects with no BMLs (group 2) and columns 4: subjects with no BMLs in medial tibia (group 3), respectively.

Feature	**Group 1**	**Group 2**	**Group 3**
Absolute Gradient			
Mean (dx)	**1.610**	1.143	0.977
Mean (dy)	**1.670**	1.173	0.910
Mean (laplacian)	**1.949**	1.138	1.381
Variance (dx)	1.112	0.885	**1.359**
Variance (dy)	**2.118**	1.529	1.078
Variance (laplacian)	**1.548**	1.066	0.937
Skewness (dx)	**2.061**	1.425	1.084
Skewness (dy)	**1.553**	1.136	0.929
Skewness (laplacian)	**1.633**	1.347	1.017
Kurtosis (dx)	**2.004**	1.417	1.155
Kurtosis (dy)	**1.591**	1.150	0.955
Kurtosis (laplacian)	**1.766**	1.326	0.996
NPWNZG(dx)	**1.940**	1.496	1.119
NPWNZG (dy)	**1.566**	1.143	0.921
NPWNZG (laplacian)	**1.781**	1.247	1.073
GLCM			
dissimilarity	**1.697**	1.022	1.054
correlation	**1.916**	1.135	1.179
contrast	**1.528**	1.044	1.133
homogeneity	**1.750**	1.346	1.122
ASM	**1.924**	1.272	1.220
energy	**1.990**	1.268	1.362
Histogram			
Mean	**1.766**	1.055	0.899
Variance	**1.270**	1.246	1.161
Skewness	1.996	1.698	**2.552**
Kurtosis	**2.044**	1.582	2.023
Fractal Dimension			
FD_hor_scale 1	1.078	0.970	**1.282**
FD_hor_scale 2	**3.988**	1.239	0.933
FD_ver_scale 1	**1.670**	1.594	1.293
FD_ver_scale 2	**4.123**	0.692	1.667

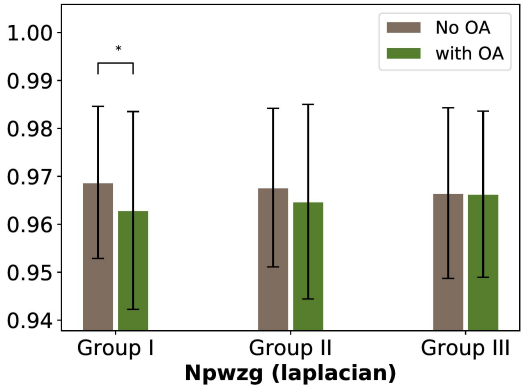


(a)

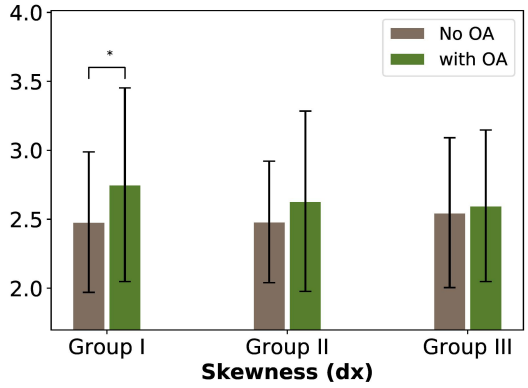
(b)



(c)



(d)



(e)

Figure 14. Mean and 95% confidence intervals of gradient-based features. Computed in relation to OA/non-OA and for each subject group independently. group 1: subjects with BMLs in medial tibia. group 2: subjects without BMLs. group 3: subjects without BMLs in medial tibia (group3). * $p < 0.05$

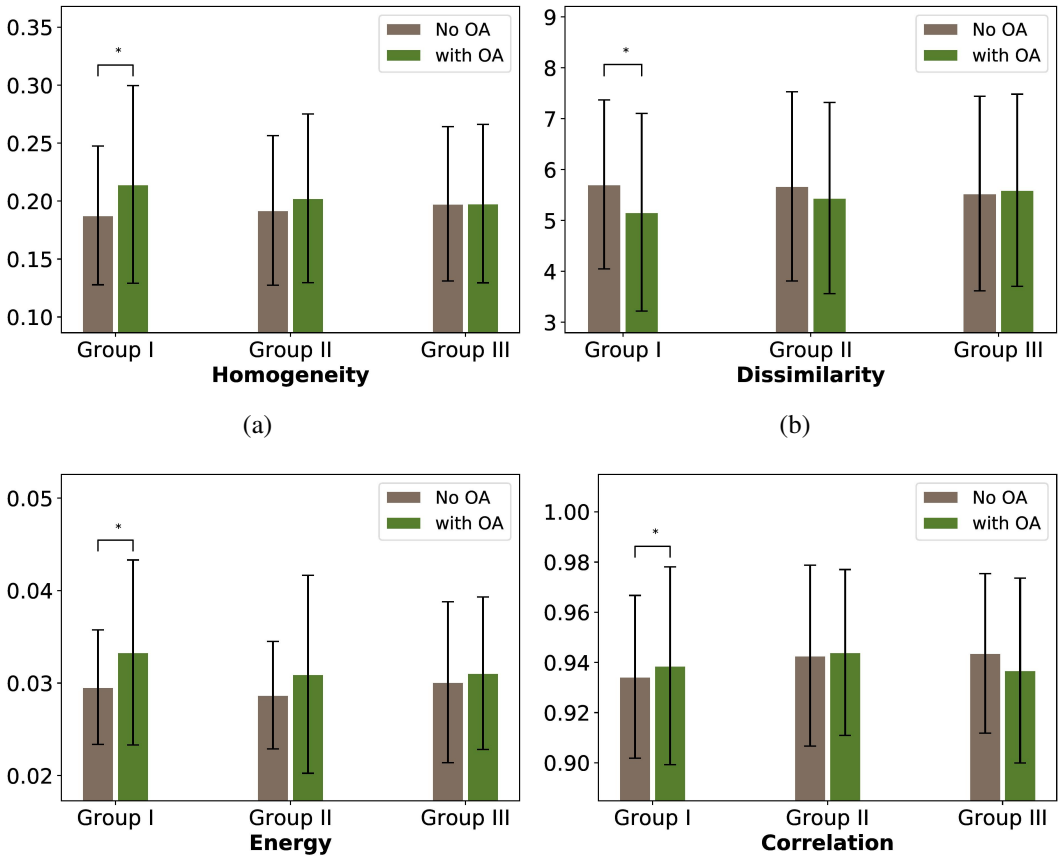


Figure 15. Mean and 95% confidence intervals of GLCM-based features. Computed in relation to OA/non-OA and for each subject group independently. group 1: subjects with BMLs in medial tibia. group 2: subjects without BMLs. group 3: subjects without BMLs in medial tibia. * $p < 0.05$

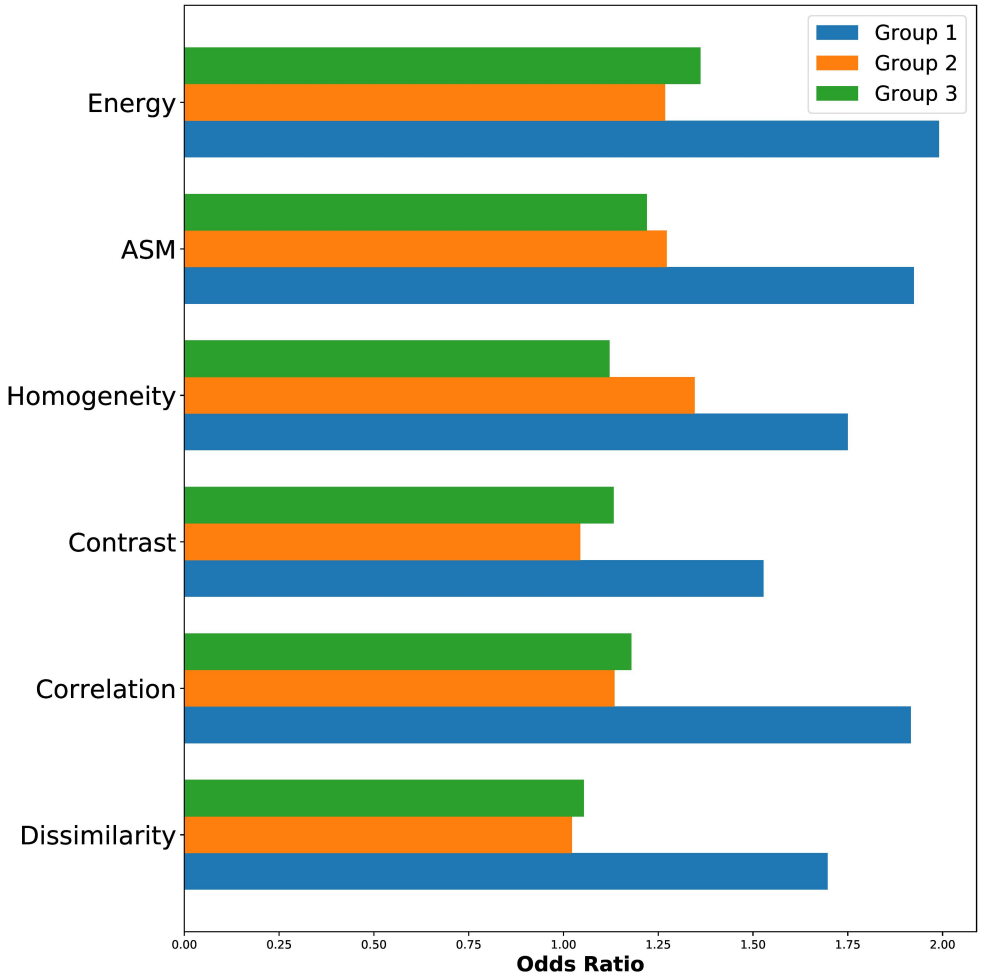


Figure 16. Comparison of odds ratio (OR) obtained from binary classification (OA vs. non OA) between gray-level co-occurrence matrix (GLCM) features of group 1 (subjects with bone marrow lesions (BML) in medial tibia), group 2 (subjects without BMLs), and group 3 (subjects without BMLs in medial tibia)

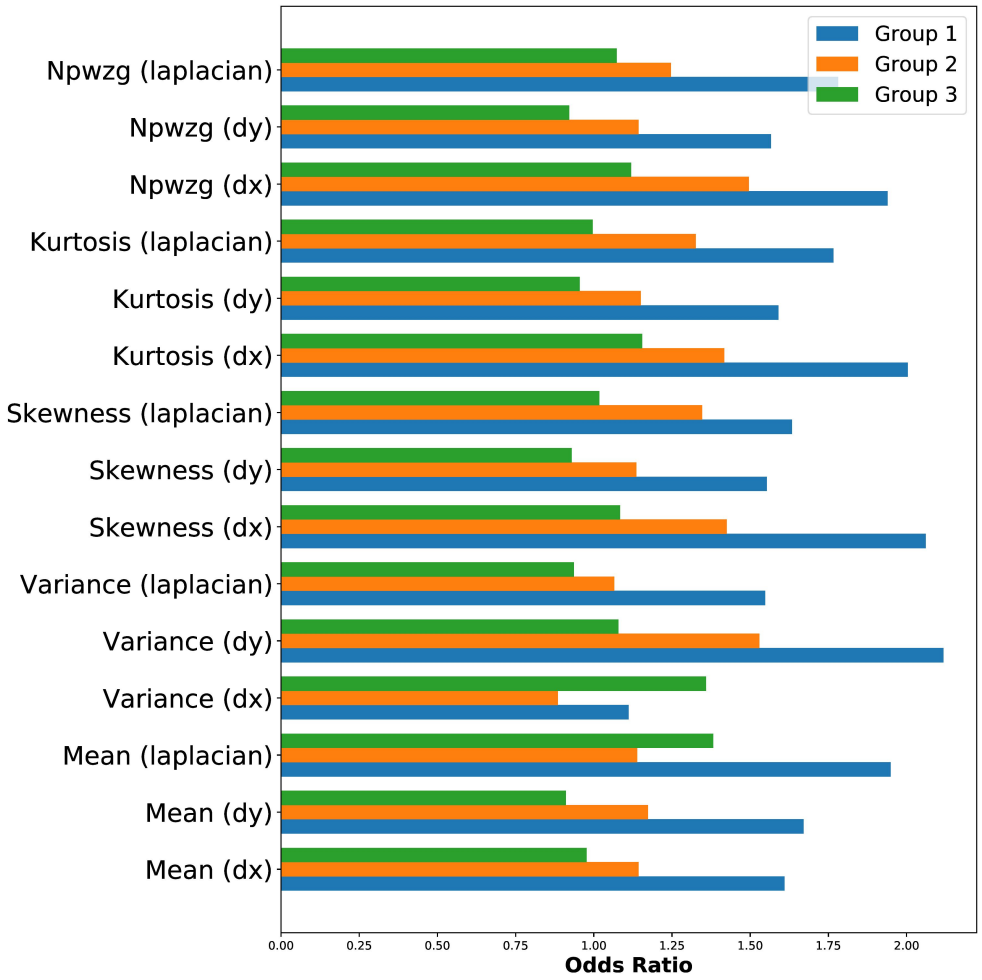


Figure 17. Comparison of odds ratio (OR) obtained from binary classification (OA vs. non OA) between absolute gradient features of group 1 (subjects with BMLs in medial tibia), group 2 (subjects without BMLs), and group 3 (subjects without BMLs in medial tibia)

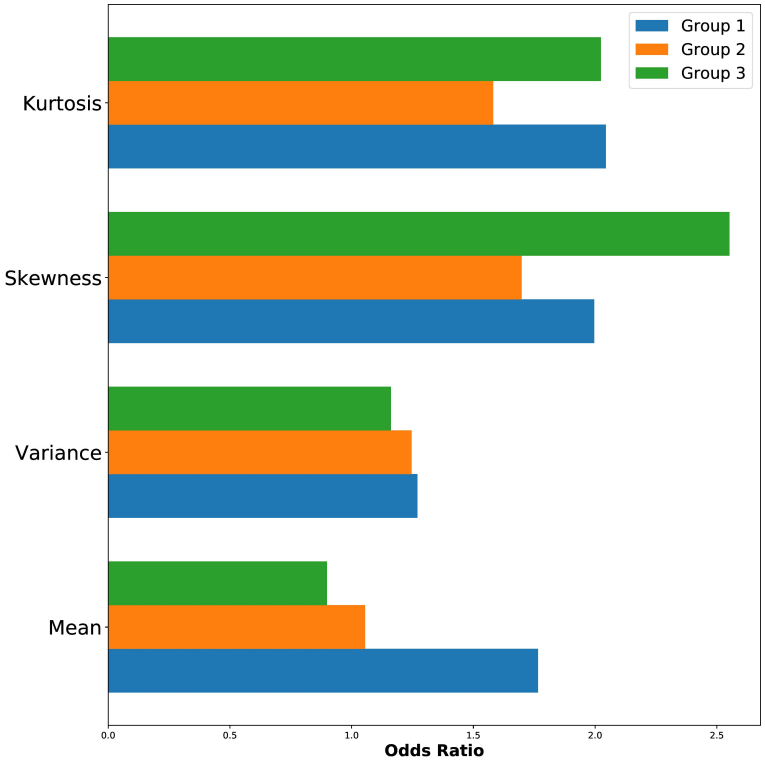


Figure 18. Comparison of odds ratio (OR) obtained from binary classification (OA vs. non OA) between histogram features of group 1 (subjects with BMLs in medial tibia), group 2 (subjects without BMLs), and group 3 (subjects without BMLs in medial tibia)

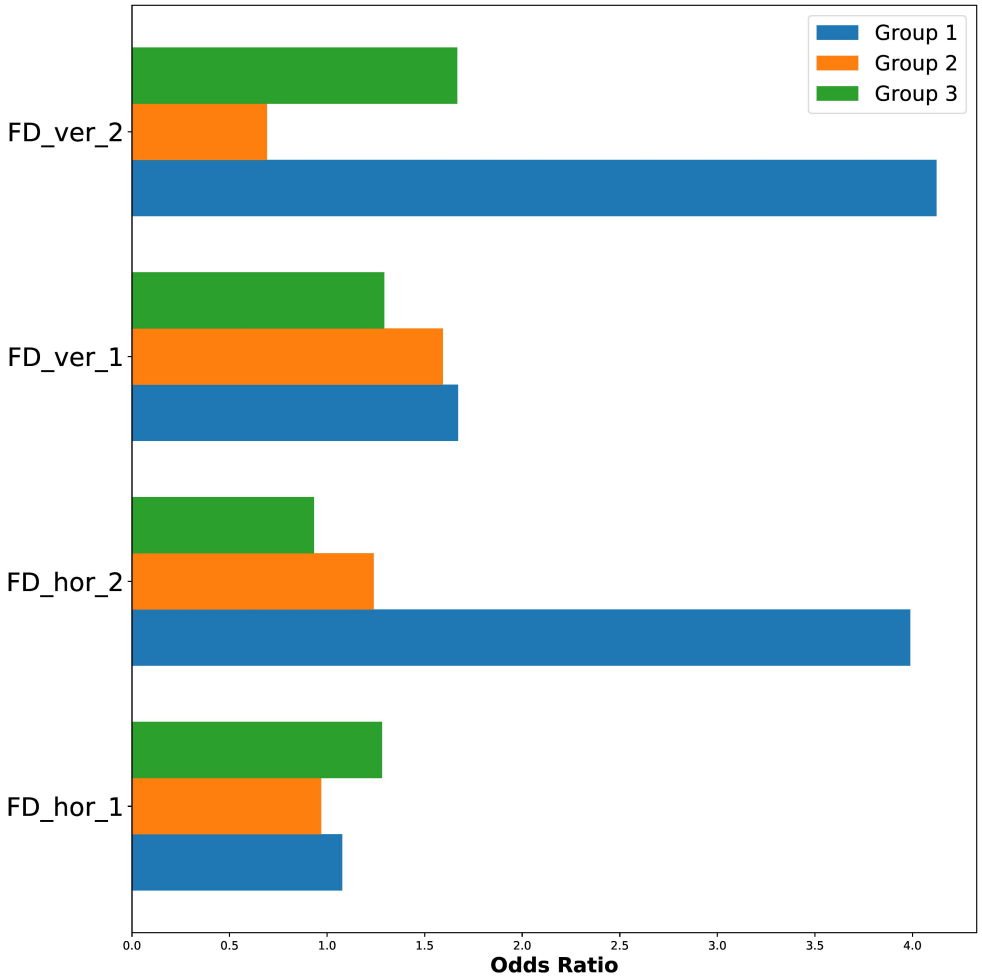


Figure 19. Comparison of odds ratio (OR) obtained from binary classification (OA vs. non OA) between fractal dimension (FD) features of group 1 (subjects with BMLs in medial tibia), group 2 (subjects without BMLs), and group 3 (subjects without BMLs in medial tibia)

Table 9. p values of Mann-Whitney U Test for each textural feature among subjects with BML at medial tibia (group 1), subjects without BML (group 2) and subjects without BMLs at medial tibia (group 3)

Mann-Whitney U Test			
Feature	p (Group 1 vs. Group 2)	p (Group 1 vs. Group 3)	p (Group 2 vs. Group 3)
Absolute Gradient			
Mean(dx)	**0.005**	0.103	0.100
Mean(dy)	**<0.001**	**<0.001**	0.383
Mean(laplacian)	**<0.001**	**<0.001**	0.337
Variance(dx)	0.474	0.359	0.369
Variance(dy)	**<0.001**	**<0.001**	0.297
Variance(laplacian)	**<0.001**	**0.001**	0.151
Skewness(dx)	**<0.001**	**0.002**	**0.016**
Skewness(dy)	0.090	0.405	0.136
Skewness(laplacian)	**<0.001**	**0.005**	0.118
Kurtosis(dx)	**<0.001**	**0.001**	**0.039**
Kurtosis(dy)	0.194	0.241	**0.050**
Kurtosis(laplacian)	**<0.001**	**0.004**	0.119
NPWNZG(dx)	**<0.001**	**0.005**	0.116
NPWNZG(dy)	**0.007**	0.070	0.185
NPWNZG(laplacian)	**<0.001**	**0.018**	0.117
GLCM			
Dissimilarity	**<0.001**	**0.001**	0.290
Correlation	**0.009**	0.386	**0.016**
Contrast	**<0.001**	**<0.001**	0.483
Homogeneity	**<0.001**	**0.008**	0.121
ASM[3]	**<0.001**	**<0.001**	**0.001**
Energy	**<0.001**	**<0.001**	**0.001**
Histogram			
Mean	**<0.001**	**<0.001**	0.07
Variance	**<0.001**	**<0.001**	**<0.001**
Skewness	**<0.001**	**<0.001**	**<0.001**
Kurtosis	**<0.001**	**<0.001**	**<0.001**

4. DISCUSSION

In this work, we analysed subchondral bone texture from 2-D plain radiographs in order to investigate the association of BMLs with OA diagnosis. Our findings demonstrate that the presence of BMLs can lead to significant changes at subchondral texture. Moreover, it may cause significant textural differences between OA and non OA subjects. Besides, our results indicated that BMLs may be locally associated with textural changes of subchondral bone. Meaning that the presences of BMLs in particular sub-regions probably cannot affect bone texture in other sub-regions in tibia (or femur).

There were no sufficient evidence to imply that alterations of subchondral bone structure observed in radiographic images directly arose from BMLs. However, recent studies indicated that subchondral bone textures are different between subjects with and without BMLs from plain radiographs and MR images [62, 58]. It means that the presence of BMLs as well as the presence of OA, is associated with textural changes in subchondral bone [60]. Nevertheless, the effect of BMLs on knee OA diagnosis based on the subchondral bone texture, has not been fully investigated yet.

We discovered that individuals with and without OA have the most significant differences in the textural features when they have BMLs in medial tibia (see Figure 14 and Figure 15). Possible explanation is that by taking BMLs into account, textural changes in subchondral bone between subjects with and without OA became more visible or distinguishable and this will help diagnosing knee OA from plain radiographs. Additionally, there are no noticeable differences in subchondral bone texture between subjects without BMLs and subjects without BMLs in medial tibia. The number of features which have significant differences between OA and non OA in both groups were considerably low compared to the subjects with BMLs in medial tibia. Comparisons between these two groups can raise a question of whether location of BMLs is an influential factor that should be considered in texture analysis of subchondral bone in OA studies. Subjects with no-BML at medial tibia included individuals whom have BMLs in sub-spinous or lateral sub-regions of tibia. One interpretation is that the presence of BML in each sub-region may presumably effects on textural features in the same sub-regions. Otherwise, these two groups would have achieved more significant differences in textural features than subjects with no-BML. Nevertheless, further study needs to be done to support this hypothesis. In fact, since we consider binary BML score for this work, the chances might be high that one group might have higher BML scores than the other groups. Thus, the reliability of statistical analysis results will be enhanced if we consider multi-score BMLs instead of binary.

In this study, we showed that features from GLCM, absolute gradient, and FD were able to distinguish the differences between these subject groups but histogram features had less significant differences between OA and non OA compared to other descriptors. Merely mean and variance out of total 4 features from histogram showed differences between the groups. On the contrary, GLCM for group 1 had 5 features (of total 6 features) had significant differences among OA and non OA while it failed to distinguish OA and non OA subjects for group 2 and group 3. This reveals that GLCM is able to catch textural differences between OA and non OA subjects when the differences are associated with BMLs. After GLCM, absolute gradient and FD features had the best performance in differentiation of subjects with OA and without OA.

One and probably the most important limitation of our study is that the subject groups were not adjusted for KL grade, gender, age, and body mass index. Adjusting such terms among subjects could establish more fair comparisons. Aside from that, the number of subjects were different between groups and also the ratio between OA and no OA classes were not same for all groups. Although we used the tests which are designed for imbalanced data, these unadjusted terms may have potential to change the results.

In future studies, employing same statistical analysis on MR images can be studied. Although, this suggestion seems convenient and straightforward, it has major challenges that we did not undergo in this thesis. For instance, it should be considered how to deal with the location of BMLs in MR images when using 2-D slices. In addition, automation of ROI extraction from MR images should be considered since it can bypass the time-consuming procedure of manual segmentation or annotation. Another future work can have one or multiple MOAKS features in textural analysis unlike this thesis which used only BML. In section 1.6, we have mentioned two papers [62, 58] that their findings indicated the effect of cartilage damage, besides BMLs, on subchondral bone texture. Perhaps, taking other features into account give us more understanding of subchondral bone alternation and its relation to OA.

In conclusion, we analysed textural features of subchondral bone for OA diagnosis and our results suggested that presence of BMLs can create detectable textural differences in subchondral bone between subjects with and without OA. However, it is crucial to consider the location of BMLs as well as its presence. Meaning that textural differences in medial tibia between OA and non-OA can be identified only when medial tibia has non-zero BML scores. In other words, having BMLs in other sub-regions may not effect on subchondral bone texture among OA and non-OA joints.

All codes used in this thesis will be published on the research unit's GitHub page (https://github.com/akazemtarghi/OABML.git).

5. REFERENCES

[1] Woolf A.D. & Pfleger B. (2003) Burden of major musculoskeletal conditions. Bulletin of the world health organization 81, pp. 646–656.

[2] Glyn-Jones S., Palmer A., Agricola R., Price A., Vincent T., Weinans H. & Carr A. (2015) Osteoarthritis. The Lancet 386, pp. 376–387.

[3] Wallace I.J., Worthington S., Felson D.T., Jurmain R.D., Wren K.T., Maijanen H., Woods R.J. & Lieberman D.E. (2017) Knee osteoarthritis has doubled in prevalence since the mid-20th century. Proceedings of the National Academy of Sciences 114, pp. 9332–9336.

[4] Terence W. O'Neill Paul S. McCabe J.M. (2018) Update on the epidemiology, risk factors and disease outcomes of osteoarthritis. Best Practice and Research Clinical Rheumatology 32, pp. 312–326.

[5] Ferket B.S., Feldman Z., Zhou J., Oei E.H., Bierma-Zeinstra S.M. & Mazumdar M. (2017) Impact of total knee replacement practice: cost effectiveness analysis of data from the osteoarthritis initiative. bmj 356, p. j1131.

[6] Lao C., Lees D., Patel S., White D. & Lawrenson R. (2019) Geographical and ethnic differences of osteoarthritis-associated hip and knee replacement surgeries in new zealand: a population-based cross-sectional study. BMJ open 9, p. e032993.

[7] Jordan J.M., Helmick C.G., Renner J.B., Luta G., Dragomir A.D., Woodard J., Fang F., Schwartz T.A., Abbate L.M., Callahan L.F. et al. (2007) Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in african americans and caucasians: the johnston county osteoarthritis project. The Journal of rheumatology 34, pp. 172–180.

[8] Suri P., Morgenroth D.C. & Hunter D.J. (2012) Epidemiology of osteoarthritis and associated comorbidities. PM&R 4, pp. S10–S19.

[9] Wright E.A., Katz J.N., Cisternas M.G., Kessler C.L., Wagenseller A. & Losina E. (2010) Impact of knee osteoarthritis on health care resource utilization in a us population-based national sample. Medical care 48, p. 785.

[10] Bombardier C., Hawker G. & Mosher D. (2015), The impact of arthritis in canada: today and over the next 30 years. toronto: Arthritis alliance of canada; 2011.

[11] Murray C., Vos T., Lozano R., AlMazroa M.A. & Memish Z.A. (2014) Disability-adjusted life years (dalys) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the global burden of disease study 2010 (vol 380, pg 2197, 2012) .

[12] Heidari B. (2011) Knee osteoarthritis prevalence, risk factors, pathogenesis and features: Part i. Caspian journal of internal medicine 2, p. 205.

[13] Lawrence R.C., Felson D.T., Helmick C.G., Arnold L.M., Choi H., Deyo R.A., Gabriel S., Hirsch R., Hochberg M.C., Hunder G.G. et al. (2008) Estimates of the prevalence of arthritis and other rheumatic conditions in the united states: Part ii. Arthritis & Rheumatism 58, pp. 26–35.

[14] Zhang Y. & Jordan J.M. (2010) Epidemiology of osteoarthritis. Clinics in geriatric medicine 26, pp. 355–369.

[15] Losina E., Walensky R.P., Reichmann W.M., Holt H.L., Gerlovin H., Solomon D.H., Jordan J.M., Hunter D.J., Suter L.G., Weinstein A.M. et al. (2011) Impact of obesity and knee osteoarthritis on morbidity and mortality in older americans. Annals of internal medicine 154, pp. 217–226.

[16] Hunter D.J., Schofield D. & Callander E. (2014) The individual and socioeconomic impact of osteoarthritis. Nature Reviews Rheumatology 10, pp. 437–441.

[17] Gupta S., Hawker G., Laporte A., Croxford R. & Coyte P. (2005) The economic burden of disabling hip and knee osteoarthritis (oa) from the perspective of individuals living with this condition. Rheumatology 44, pp. 1531–1537.

[18] Hunter D.J. & Bierma-zeinstra S. (2019) Seminar osteoarthritis .

[19] Whittaker J.L., Toomey C.M., Woodhouse L.J., Jaremko J.L., Nettel-Aguirre A. & Emery C.A. (2018) Association between mri-defined osteoarthritis, pain, function and strength 3–10 years following knee joint injury in youth sport. British journal of sports medicine 52, pp. 934–939.

[20] Emery C.A., Whittaker J.L., Mahmoudian A., Lohmander L.S., Roos E.M., Bennell K.L., Toomey C.M., Reimer R.A., Thompson D., Ronsky J.L. et al. (2019) Establishing outcome measures in early knee osteoarthritis. Nature Reviews Rheumatology 15, pp. 438–448.

[21] O'Neill T.W., McCabe P.S. & McBeth J. (2018) Update on the epidemiology, risk factors and disease outcomes of osteoarthritis. Best Practice & Research Clinical Rheumatology 32, pp. 312–326.

[22] Tiulpin A. (2016) Deep learning for knee osteoarthritis diagnosis and progression prediction from plain radiographs and clinical data .

[23] Mora J.C., Przkora R. & Cruz-Almeida Y. (2018) Knee osteoarthritis: pathophysiology and current treatment modalities. Journal of pain research 11, p. 2189.

[24] Yucesoy B., Charles L.E., Baker B. & Burchfiel C.M. (2015) Occupational and genetic risk factors for osteoarthritis: a review. Work 50, pp. 261–273.

[25] Adjei Y. Radiographic assessment of the association of upper femur geometry and texture features to hip osteoarthritis and fracture .

[26] Sharma V., Anuvat K., John L. & Davis M. (2017) Scientific american pain management-arthritis of the knee. Decker: Pain related disease states .

[27] Ayhan E., Kesmezacar H. & Akgun I. (2014) Intraarticular injections (corticosteroid, hyaluronic acid, platelet rich plasma) for the knee osteoarthritis. World journal of orthopedics 5, p. 351.

[28] Madry H., van Dijk C.N. & Mueller-Gerbl M. (2010) The basic science of the subchondral bone. Knee surgery, sports traumatology, arthroscopy 18, pp. 419–433.

[29] Burr D.B. & Gallant M.A. (2012) Bone remodelling in osteoarthritis. Nature Reviews Rheumatology 8, p. 665.

[30] Li G., Yin J., Gao J., Cheng T.S., Pavlos N.J., Zhang C. & Zheng M.H. (2013) Subchondral bone in osteoarthritis: insight into risk factors and microstructural changes. Arthritis research & therapy 15, p. 223.

[31] Milz S. & Putz R. (1994) Quantitative morphology of the subchondral plate of the tibial plateau. Journal of anatomy 185, p. 103.

[32] Armstrong S.J., Read R.A. & Price R. (1995) Topographical variation within the articular cartilage and subchondral bone of the normal ovine knee joint: a histological approach. Osteoarthritis and cartilage 3, pp. 25–33.

[33] Patel V., Issever A.S., Burghardt A., Laib A., Ries M. & Majumdar S. (2003) Microct evaluation of normal and osteoarthritic bone structure in human knee specimens. Journal of Orthopaedic Research 21, pp. 6–13.

[34] Clark J.M. (1990) The structure of vascular channels in the subchondral plate. Journal of anatomy 171, p. 105.

[35] Goldring M.B. & Goldring S.R. (2010) Articular cartilage and subchondral bone in the pathogenesis of osteoarthritis. Annals of the New York Academy of Sciences 1192, pp. 230–237.

[36] MacKay J.W., Kapoor G., Driban J.B., Lo G.H., McAlindon T.E., Toms A.P., McCaskie A.W. & Gilbert F.J. (2018) Association of subchondral bone texture on magnetic resonance imaging with radiographic knee osteoarthritis progression: data from the osteoarthritis initiative bone ancillary study. European radiology 28, pp. 4687–4695.

[37] Klement M.R. & Sharkey P.F. (2019) The significance of osteoarthritis-associated bone marrow lesions in the knee. JAAOS-Journal of the American Academy of Orthopaedic Surgeons 27, pp. 752–759.

[38] Eriksen E.F. (2015) Treatment of bone marrow lesions (bone marrow edema). BoneKEy reports 4.

[39] MacKay J.W., Murray P.J., Kasmai B., Johnson G., Donell S.T. & Toms A.P. (2016) Mri texture analysis of subchondral bone at the tibial plateau. European radiology 26, pp. 3034–3045.

[40] Hunter D.J., Zhang Y., Niu J., Goggins J., Amin S., LaValley M.P., Guermazi A., Genant H., Gale D. & Felson D.T. (2006) Increase in bone marrow lesions associated with cartilage loss: a longitudinal magnetic resonance imaging study of knee osteoarthritis. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology 54, pp. 1529–1535.

[41] Tanamas S.K., Wluka A.E., Pelletier J.P., Pelletier J.M., Abram F., Berry P.A., Wang Y., Jones G. & Cicuttini F.M. (2010) Bone marrow lesions in people with knee osteoarthritis predict progression of disease and joint replacement: a longitudinal study. Rheumatology 49, pp. 2413–2419.

[42] Hunter D.J., Guermazi A., Lo G.H., Grainger A.J., Conaghan P.G., Boudreau R.M. & Roemer F.W. (2011) Evolution of semi-quantitative whole joint assessment of knee oa: Moaks (mri osteoarthritis knee score). Osteoarthritis and cartilage 19, pp. 990–1002.

[43] Sakellariou G., Conaghan P.G., Zhang W., Bijlsma J.W., Boyesen P., D'Agostino M.A., Doherty M., Fodor D., Kloppenburg M., Miese F. et al. (2017) Eular recommendations for the use of imaging in the clinical management of peripheral joint osteoarthritis. Annals of the rheumatic diseases 76, pp. 1484–1494.

[44] Hunter D.J., McDougall J.J. & Keefe F.J. (2008) The symptoms of osteoarthritis and the genesis of pain. Rheumatic Disease Clinics of North America 34, pp. 623–643.

[45] Hirvasniemi J. (2015) Novel x-ray-based methods for diagnostics of osteoarthritis. University Of Oulu, Institute of Clinical Medicine and Department of Applied Physics .

[46] Braun H.J. & Gold G.E. (2012) Diagnosis of osteoarthritis: imaging. Bone 51, pp. 278–288.

[47] Roemer F.W., Jarraya M., Felson D.T., Hayashi D., Crema M., Loeuille D. & Guermazi A. (2016) Magnetic resonance imaging of hoffa's fat pad and relevance for osteoarthritis research: a narrative review. Osteoarthritis and cartilage 24, pp. 383–397.

[48] Podlipská J., Guermazi A., Lehenkari P., Niinimäki J., Roemer F.W., Arokoski J.P., Kaukinen P., Liukkonen E., Lammentausta E., Nieminen M.T. et al. (2016) Comparison of diagnostic performance of semi-quantitative knee ultrasound and knee radiography with mri: Oulu knee osteoarthritis study. Scientific reports 6, pp. 1–12.

[49] Roemer F.W., Crema M.D., Trattnig S. & Guermazi A. (2011) Advances in imaging of osteoarthritis and cartilage. Radiology 260, pp. 332–354.

[50] Kellgren J. & Lawrence J. (1957) Radiological assessment of osteo-arthrosis. Annals of the rheumatic diseases 16, p. 494.

[51] Felson D.T., Naimark A., Anderson J., Kazis L., Castelli W. & Meenan R.F. (1987) The prevalence of knee osteoarthritis in the elderly. the framingham osteoarthritis study. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology 30, pp. 914–918.

[52] Kohn M.D., Sassoon A.A. & Fernando N.D. (2016), Classifications in brief: Kellgren-lawrence classification of osteoarthritis.

[53] Peterfy C., Guermazi A., Zaim S., Tirman P., Miaux Y., White D., Kothari M., Lu Y., Fye K., Zhao S. et al. (2004) Whole-organ magnetic resonance imaging

score (worms) of the knee in osteoarthritis. Osteoarthritis and cartilage 12, pp. 177–190.

[54] Hunter D.J., Lo G.H., Gale D., Grainger A.J., Guermazi A. & Conaghan P.G. (2008) The reliability of a new scoring system for knee osteoarthritis mri and the validity of bone marrow lesion assessment: Bloks (boston–leeds osteoarthritis knee score). Annals of the rheumatic diseases 67, pp. 206–211.

[55] Stewart H.L. & Kawcak C.E. (2018) The importance of subchondral bone in the pathophysiology of osteoarthritis. Frontiers in veterinary science 5, p. 178.

[56] Alliston T., Hernandez C.J., Findlay D.M., Felson D.T. & Kennedy O.D. (2018) Bone marrow lesions in osteoarthritis: What lies beneath. Journal of Orthopaedic Research® 36, pp. 1818–1825.

[57] Hirvasniemi J., Thevenot J., Immonen V., Liikavainio T., Pulkkinen P., Jämsä T., Arokoski J. & Saarakkala S. (2014) Quantification of differences in bone texture from plain radiographs in knees with and without osteoarthritis. Osteoarthritis and cartilage 22, pp. 1724–1731.

[58] Wolski M., Stachowiak G.W., Dempsey A.R., Mills P.M., Cicuttini F.M., Wang Y., Stoffel K.K., Lloyd D.G. & Podsiadlo P. (2011) Trabecular bone texture detected by plain radiography and variance orientation transform method is different between knees with and without cartilage defects. Journal of Orthopaedic Research 29, pp. 1161–1167.

[59] Bayramoglu N., Tiulpin A., Hirvasniemi J., Nieminen M.T. & Saarakkala S. (2020) Adaptive segmentation of knee radiographs for selecting the optimal roi in texture analysis. Osteoarthritis and Cartilage .

[60] Hirvasniemi J., Niinimäki J., Thevenot J. & Saarakkala S. (2019) Bone density and texture from minimally post-processed knee radiographs in subjects with knee osteoarthritis. Annals of biomedical engineering 47, pp. 1181–1190.

[61] Janvier T., Jennane R., Valery A., Harrar K., Delplanque M., Lelong C., Loeuille D., Toumi H. & Lespessailles E. (2017) Subchondral tibial bone texture analysis predicts knee osteoarthritis progression: data from the osteoarthritis initiative: Tibial bone texture & knee oa progression. Osteoarthritis and cartilage 25, pp. 259–266.

[62] Hirvasniemi J., Thevenot J., Guermazi A., Podlipská J., Roemer F.W., Nieminen M.T. & Saarakkala S. (2017) Differences in tibial subchondral bone structure

evaluated using plain radiographs between knees with and without cartilage damage or bone marrow lesions-the oulu knee osteoarthritis study. European radiology 27, pp. 4874–4882.

[63] Wise B.L., Niu J., Guermazi A., Liu F., Heilmeier U., Ku E., Lynch J.A., Zhang Y., Felson D.T., Kwoh C.K. et al. (2017) Magnetic resonance imaging lesions are more severe and cartilage t2 relaxation time measurements are higher in isolated lateral compartment radiographic knee osteoarthritis than in isolated medial compartment disease–data from the osteoarthritis initiative. Osteoarthritis and cartilage 25, pp. 85–93.

[64] Lindner C., Thiagarajah S., Wilkinson J.M., Wallis G.A., Cootes T.F., arcOGEN Consortium et al. (2013) Accurate bone segmentation in 2d radiographs using fully automatic shape model matching based on regression-voting. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp. 181–189.

[65] Haralick R.M., Shanmugam K. & Dinstein I.H. (1973) Textural features for image classification. IEEE Transactions on systems, man, and cybernetics , pp. 610–621.

[66] Ojala T., Pietikainen M. & Maenpaa T. (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence 24, pp. 971–987.

[67] Bharati M.H., Liu J.J. & MacGregor J.F. (2004) Image texture analysis: methods and comparisons. Chemometrics and intelligent laboratory systems 72, pp. 57–71.

[68] Zhang J. & Tan T. (2002) Brief review of invariant texture analysis methods. Pattern recognition 35, pp. 735–747.

[69] Tou J.Y., Tay Y.H. & Lau P.Y. (2009) Recent trends in texture classification: a review. In: Symposium on Progress in Information & Communication Technology, vol. 3, vol. 3, pp. 56–59.

[70] Mehri M., Héroux P., Gomez-Krämer P. & Mullot R. (2017) Texture feature benchmarking and evaluation for historical document image analysis. International Journal on Document Analysis and Recognition (IJDAR) 20, pp. 1–35.

[71] Feng D.D. (2011) Biomedical information technology. Academic Press.

[72] Buckland-Wright J., Lynch J. & Macfarlane D. (1996) Fractal signature analysis measures cancellous bone organisation in macroradiographs of patients with knee osteoarthritis. Annals of the rheumatic diseases 55, pp. 749–755.

[73] Podsiadlo P., Nevitt M., Wolski M., Stachowiak G., Lynch J., Tolstykh I., Felson D., Segal N., Lewis C. & Englund M. (2016) Baseline trabecular bone and its relation to incident radiographic knee osteoarthritis and increase in joint space narrowing score: directional fractal signature analysis in the most study. Osteoarthritis and cartilage 24, pp. 1736–1744.

[74] Hirvasniemi J., Thevenot J., Multanen J., Haapea M., Heinonen A., Nieminen M.T. & Saarakkala S. (2017) Association between radiography-based subchondral bone structure and mri-based cartilage composition in postmenopausal women with mild osteoarthritis. Osteoarthritis and cartilage 25, pp. 2039–2046.

[75] Harrar K. & Jennane R. (2015) Trabecular texture analysis using fractal metrics for bone fragility assessment. structure 5, p. 6.

[76] (2018 (accessed February, 2020)), developed by german cancer research center division of medical image computing. `http://mitk.org/wiki/MITK`.

[77] Dutta A., Gupta A. & Zissermann A. (2016), VGG image annotator (VIA). http://www.robots.ox.ac.uk/ vgg/software/via/.

[78] Dutta A. & Zisserman A. (2019) The VIA annotation software for images, audio and video. In: Proceedings of the 27th ACM International Conference on Multimedia, MM '19, ACM, New York, NY, USA.

[79] Lynch J., Hawkes D. & Buckland-Wright J. (1991) Analysis of texture in macroradiographs of osteoarthritic knees, using the fractal signature. Physics in Medicine & Biology 36, p. 709.

[80] Thomson J., O'Neill T., Felson D. & Cootes T. (2015) Automated shape and texture analysis for detection of osteoarthritis from radiographs of the knee. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp. 127–134.

[81] Hafezi-Nejad N., Guermazi A., Demehri S. & Roemer F.W. (2017) New imaging modalities to predict and evaluate osteoarthritis progression. Best Practice & Research Clinical Rheumatology 31, pp. 688–704.

6. APPENDIX A

6.1. Sub-Study Overview

In the following, the second sub-study is described. The aim of this sub-study is to investigate whether our convolutional neural network (CNN) model is able to detect OA by using subchondral bone texture when subjects have no bone marrow lesions (BML). In the first section, we described our subject selection. Then preprocessing methods were explained. Subsequently, information related to implementation and our proposed CNN network was provided. At the end, the results were reported.

6.2. Subject Selection

Study Subjects for MR images: We selected the subjects who had available SAG 3D DESS WE sequence besides plain radiography with no BML (0 BML score according to MRI osteoarthritis knee score (MOAKS)) in both femur and tibia at the baseline. Subjects with missing assessment information were excluded. After exclusions, there were 658 images from 613 subjects (some subjects have two images from the both knees). Among them, 123 and 535 images were with and without OA, respectively. In order to have more balance data set, we randomly eliminated nearly half of images without OA. Finally, we had 250 images without OA and 123 with OA (see Figure A.1).

6.3. Preprocessing of 3-D DESS MR Images

Preprocessing steps for MRI data are 1) selection of the most central coronal slice for each subject 2) manual annotation of the bone contour and 3) extracting the ROI (Figure A.2).

1. **Slice Selection:** slices which show the most central part of the tibia and femur were selected. Manual selection was adopted because the most central slice of each subject was different compared to others. Since, in sagittal view, femur and tibia are not aligned vertically, we select two slices, one for central tibia and another for central femur. The choosing the central slice of each bone was challenging due to different shape of these bones from a subject to another. For manual selection of central slice, We used medical image viewer named

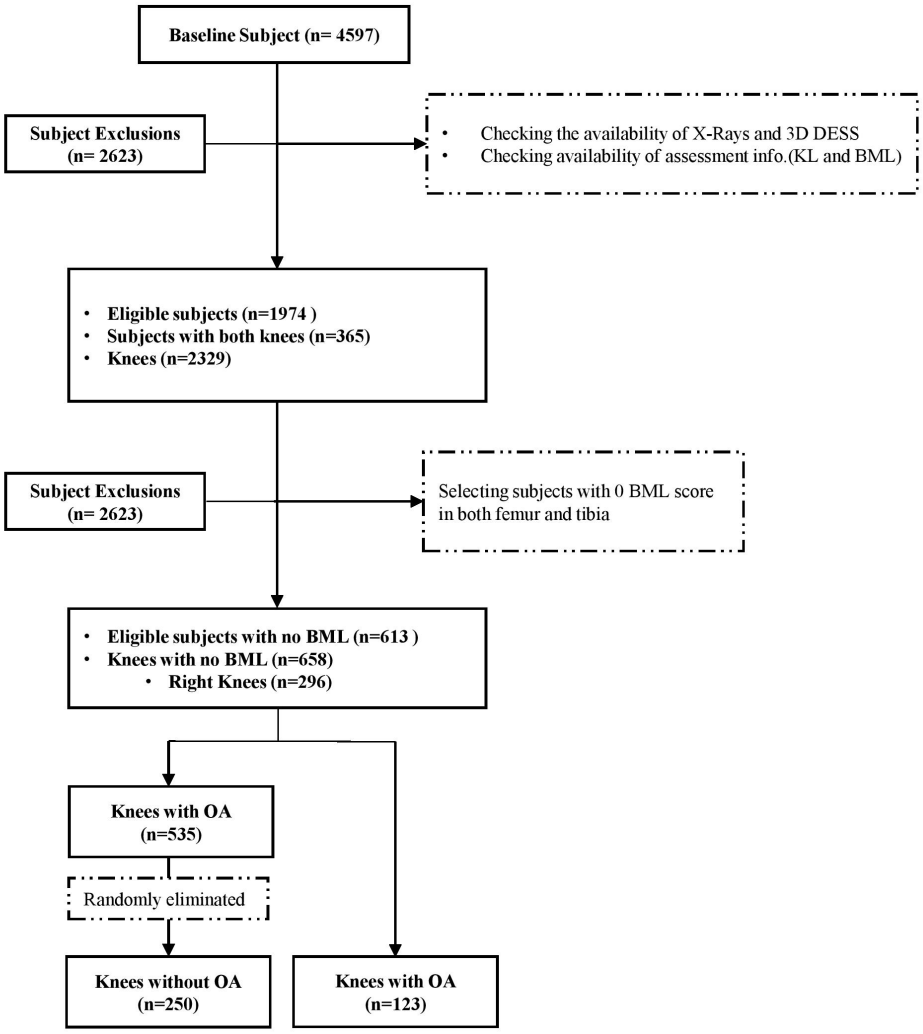


Figure A.1. The flow-gram of subject selection for deep learning-based study

The Medical Imaging Interaction Toolkit (MITK) developed by German Cancer Research Center, Heidelberg, Germany [76].

2. **Manual Annotation:** For landmark annotation, VGG Image Annotator developed by Department of Engineering Science, University of Oxford, England, was used [77, 78]. Annotation was performed according to BoneFinder landmarks for plain radiographs in which 18 from total landmarks for tibia were chosen (see Figure A.3).
3. **ROI Extraction:** The marginal ROI could not be used for MR images considering the fact that extracted ROIs most likely might contain soft tissues. Therefore, for MR images, we selected the standard ROI patches used in several

OA studies [59, 79, 62, 80, 81, 58]. In this ROI selection method, square patches with sizes proportional to the width between marginal medial tibia point and the centre of the medial condyle were localized beneath the tibial plateau (see Figure A.4). Same ROI was used for plain radiographs

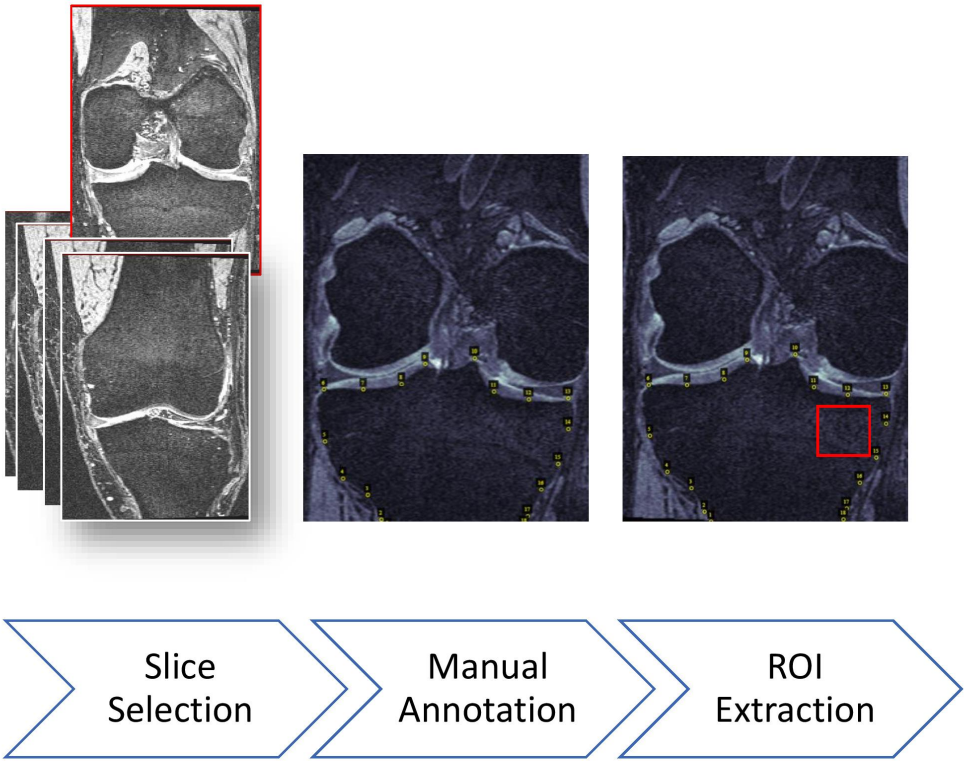


Figure A.2. The pipeline of magnetic resonance imaging (MRI) data preprocessing. The first step was selecting the most central slice, then annotating the landmarks and in the last step, region of interest (ROI) in medial tibia was extracted

6.4. CNN Architecture and Implementation

Network: In this study, deep CNN models are used to extract textural features. Our proposed network contains four convolutional layers followed by one fully connected layer (see Figure A.5). We considered a small architecture with minimal layers.

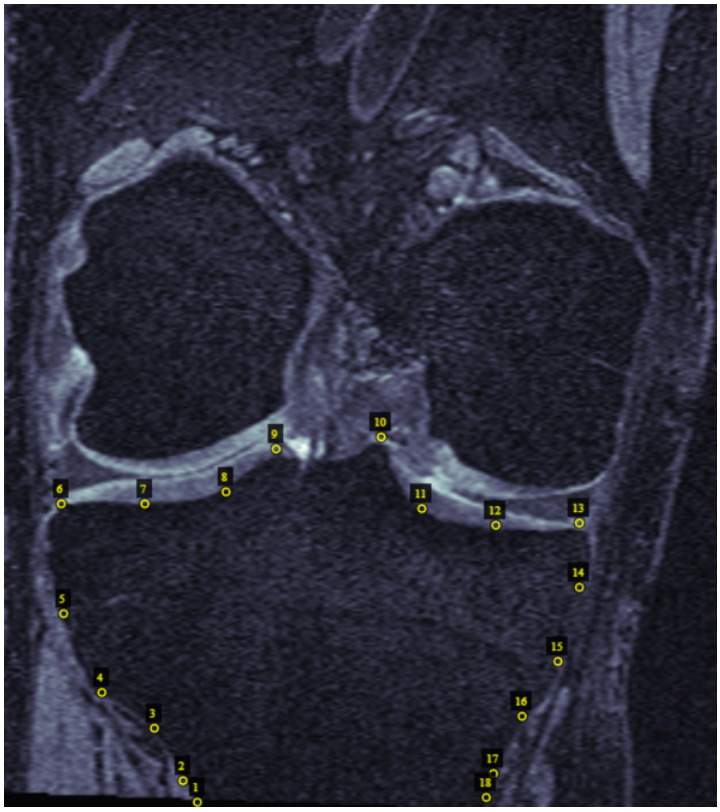


Figure A.3. One example of annotated MR image. 18 landmarks were selected on the edge of tibia. The annotation was perfromed by VGG Image Annotator [77, 78].

The network uses a stack of 3x3 convolutional filters. Each of them is followed by hyperbolic tangent (tanh) as a non-linear activation function. In the second layer, we used 3x3 max-pooling with a convolution stride of 2 where each one reduced computations. Last layer of our network contains two fully connected with two tanh activation function and two outputs OA and non OA. For regularisation, one dropout layer with ratio of 0.25 was used to avoid over-fitting.

Implementation: The sizes of our input imageswere not fixed since the standard ROIs was proportional to the size of the bone. Re-scaling the images to the same size could lead to the distortion of the subchondral bone textures. Instead, we selected fixed-size square from the centre of the ROIs to maintain the original textures. Fixed patches from input images were cropped and considered as the new inputs.

The dataset was divided into training, validation and testing parts with a rate of 14:3:3 (70 % , 15 % , 15 %). In order to implement binary classification, the data were categorised to OA (with KL grade of 1 and 2) and non OA (with KL grade of 2, 3, 4) classes. In our experiments, we used stochastic gradient descent optimiser and a cross-entropy loss function. We optimised the network with the fixed learning rate of $1e^{-3}$.

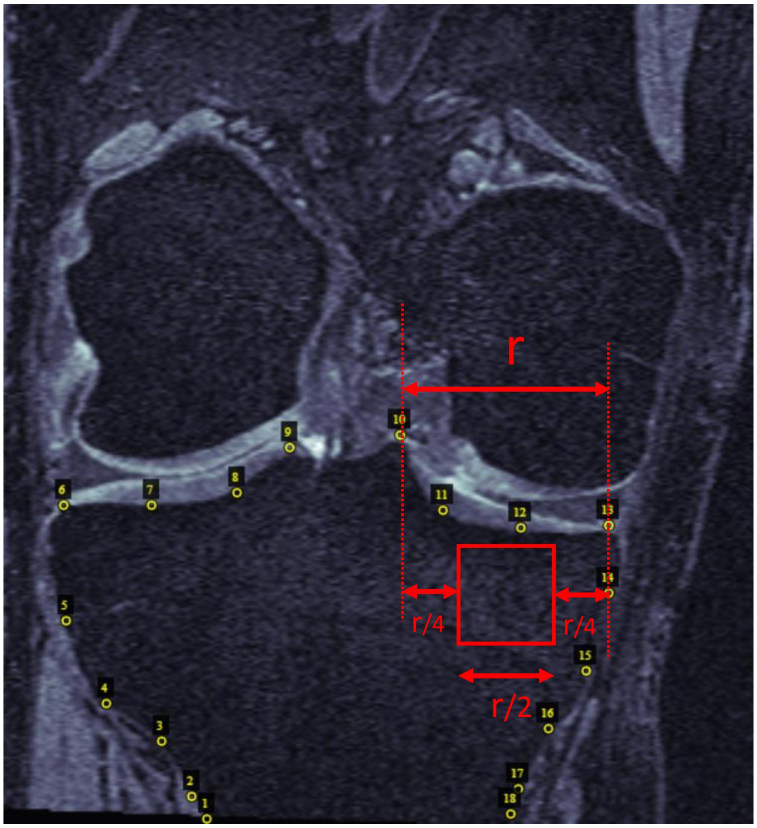


Figure A.4. Region of interest (ROI) used for MR images. The size of ROI is proportional to the size of the bone (the width between margin of tibia point and the centre of the medial condyle)

A stratified 5-fold cross-validation based on the number of labels (OA and non OA) was applied in order to obtain better analysis over the training dataset since it provides a range of results across the dataset. In stratified K-fold cross-validation, the training data is divided into K equal segments but the ratio of samples from each class is preserved in each segment. Then K-1 segments are trained and the remaining one is considered as the predictive performance which is, in other words, the expected loss on unseen future data. This process happens K times, using different test partitions. The results would be the average of outputs of all folds.

The area under the receiver operating characteristic curve (AUC-ROC) which is also called c-index and normalised confusion matrix was chosen as assessment tools of our algorithm performance. It is a common and effective method for evaluating the performance of the classification learning algorithm since ROC is representing the probability curve and AUC is a measure of separability.

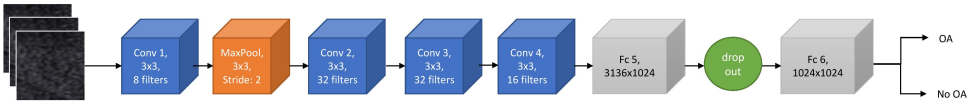


Figure A.5. The convolutional neural networks (CNN) architecture composing 6 layers, 4 convolutional followed by two fully connected layer. The kernel size of convolutional filter is 3x3.

6.5. Results

We present ROC AUC values and confusion matrices of binary classification from MR images and plain radiographs (Figure A.6). According to confusion matrix, all subjects were wrongly predicted as non OA subjects. One explanation can be the textual features might be not robust enough to discriminate the OA and non OA subjects. This hypothesis can be supported by ROC AUC which show that the model was failed to learn the textural features between the non OA subjects from OA subjects. Although variety of regularisation methods were used in order to avoid overfitting, the performance of the model did not improve.

Before experiment, we expected that the model for plain radiographs would have better performance than MRI because plain radiograph is a projection image of 3-D object and contain much more structures than single slice of MR image. However, the results from both modalities are almost similar. No differences between results of MRI and plain radiography raise a question of whether textural features between OA and non OA subjects who have no BMLs, are significantly discriminant. However, it should be noted that limitations such as the low number of data (373 knee images) may have potential to affect the results.

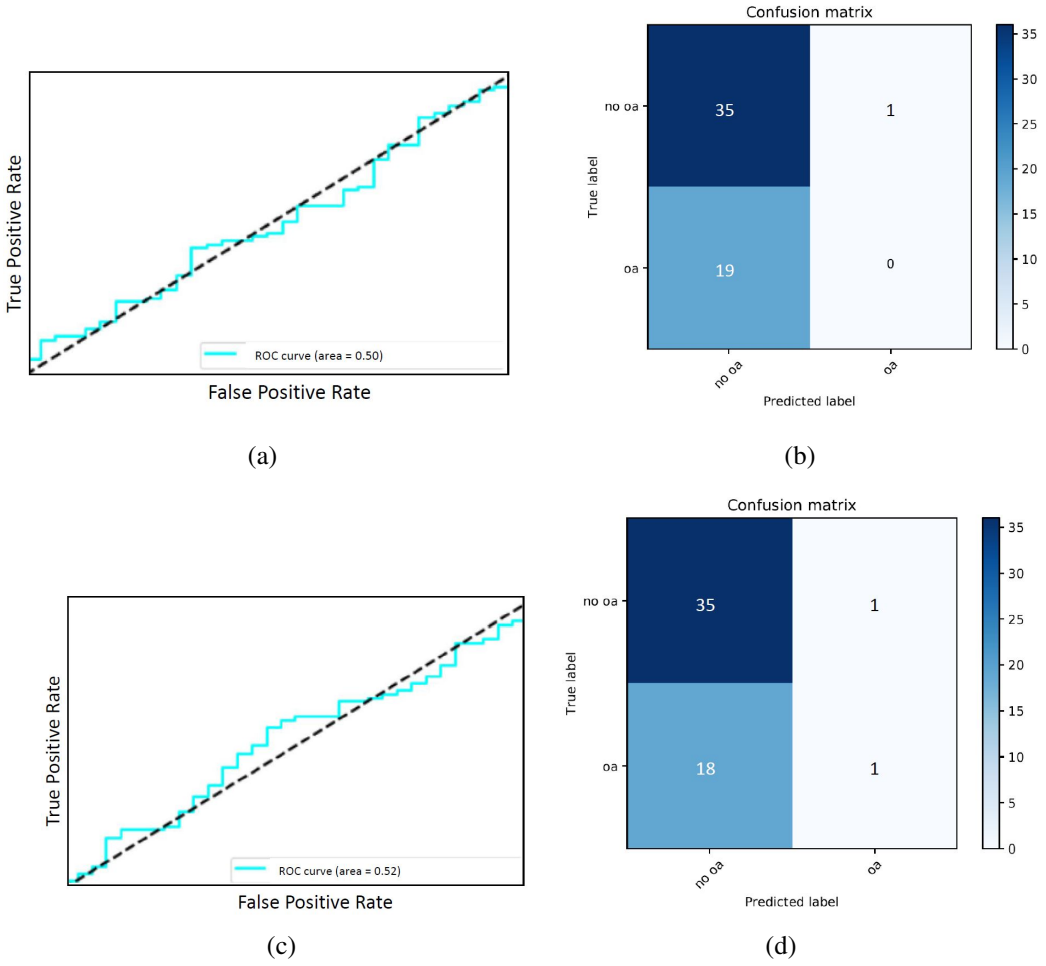


Figure A.6. Confusion matrix and Area Under the Receiver Operating Characteristic curve (AUC ROC) from binary classification. Results from MRI data (a) and (b), and from plain radio graphs (c) and (d).