PhD thesis

Joselene Marques

Osteoarthritis Imaging by Quantification of Tibial Trabecular Bone

Preface

The work described in this dissertation was carried out at eScience Center at the Department of Computer Science of the University of Copenhagen between January 2010 and December 2012 in partial fulfillment of the requirements for the degree of doctor of philosophy.

The PhD research took place at Biomediq under the supervision of professor Mads Nielsen and co-supervision of Erik B. Dam. Five weeks of this research were carried out at the Pattern Recognition Laboratory in Delft University of Technology in collaboration with Prof. Marco Loog and Prof. David Tax.

This dissertation is based on research papers published or in review in international journals and conferences of medical imaging and osteoarthritis.

Abstract

The pathogenesis of osteoarthritis (OA) includes complex events in the whole joint. Cartilage loss and bone remodelling are central in OA progression. In this project, we investigated the feasibility of quantifying OA by analysis of the tibial trabecular bone structure in low-field knee magnetic resonance imaging (MRI). The development of automatic and more sensitive indicators of OA in conjunction with low cost equipment have the potential to decrease the length and cost of clinical trials. We present a texture analysis methodology that combined machine-learning techniques in a fully automatic framework. Different linear feature selection approaches where investigated. The methodology was evaluated in a longitudinal study, where MRI scans of knees were used to quantify the tibial trabecular bone in a bone marker for OA diagnosis and another marker for prediction of tibial cartilage loss. The healthy and diseased subjects were defined by the Kellgren and Lawrence index assigned by radiologists and the levels of cartilage loss were assessed by a segmentation process. A preliminary radiological reading of the knees with high and low risks of cartilage loss suggested the prognosis bone marker captured aspects of the vertical trabecularization of the tibial bone to define the prognosis of cartilage loss. We also investigated which region of the tibia provided the best prognosis for medial tibial cartilage loss. The structure of the tibial trabecular bone was divided in localized subregions in an attempt to capture the different pathological features occurring at each location. We applied multiple-instance learning, where each subregion was defined to be one instance and a bag held all instances over a full region-of-interest. The inferior part of the tibial bone was classified as the most relevant region for prognosis of cartilage loss and a preliminary radiological reading of a subset of the samples suggested the bone marker also captured the vertical trabecularization of the tibial bone to define the most relevant region. In a clinical point of view, besides presenting a bone marker able to predict disease progression and diagnostic bone marker superior to other OA biomarkers, our findings underlined the importance of the trabecular bone to the understanding of the OA pathology.

Contents

Introduction, Background and Motivation

1.1 Osteoarthritis

Osteoarthritis (OA) is a widespread and degenerative, chronic disease that affects up to 80% of the population over 65 years of age [13]. A recent report indicates that knee OA is likely to become the fourth most common cause of disability in women and the eighth most common cause in men [107]. Recent estimates suggest that total costs for arthritis, including OA, may exceed 2% of the United States gross domestic product [17]. Due to the longevity of working careers and the prevalence of OA in middle-aged persons, OA may cause a significant burden in lost time at work and early retirement [17].

The non-modifiable risk factors include gender and age whereas the modifiable risk factors include body mass index (BMI), injury/trauma, among others.

Recent prospective studies have demonstrated that obesity is a primary risk factor for knee OA [27]. Overloading the knee joints can lead to cartilage breakdown and failure of components of structural support [51].

Before 50 years of age, the prevalence of OA in most joints is higher in men than in women. After about age 50 years, women are more often affected than men [51]. Some studies have shown women have increased rates of cartilage loss and progression of knee cartilage defects than men [52, 127].

The disturbance of the mechanical axis of the leg also influences progres-

sion of knee OA. Sharma and colleagues [90] confirmed that varus alignment was associated with an increased risk of medial tibiofemoral narrowing and valgus alignment with an increased risk of lateral tibiofemoral narrowing; the greater the varus or valgus, the greater the risk of narrowing.

The pathology of OA involves multiple components of the joint in a disease process that includes progressive degradation of articular cartilage with concomitant changes in the bone underneath the cartilage, including changes in trabecular bone structure, bone marrow lesions, development of marginal outgrowths, osteophytes and an abnormal increase in density and thickness of bone (bony sclerosis). Soft-tissue structures in and around the joint are also affected. These structures include synovium, which may reveal modest inflammatory infiltrates; ligaments, which often become lax; meniscus, which may present traumatic and degenerative lesions; and bridging muscle, which becomes weak [48,51].

Typical symptoms are swelling, pain, stiffness and decreased mobility leading to an impaired quality of life. Nevertheless, many people with pathologic and radiographic evidence of OA have no symptoms [91]. Even so, only treatments of symptoms of OA are well documented so far [116]. One reason for this, may be a insufficient understanding of the disease process and whether the joint damage is reversible or not.

The investigation presented in [8] considers a point-of-no-return in the OA disease process and discuss whether different intervention strategies may only be efficacious at distinct stages of OA. The authors suggest the use of structure-modifying OA drugs, because there is a good chance for restoring the tissue turnover, specially in early OA stages, when drug interventions seems more effective.

1.2 Is OA a bone or cartilage disease?

There has been speculation for many years that bone may be the primary organ triggering OA, microtrabecular fractures and their subsequent healing could increase the stiffness of the subchondral bone. This process could transmit increased load to overlying cartilage, leading to secondary cartilage damage [53].

Radin et al. [118] was the first to demonstrate subchondral bone changes

in OA patients and to propose that they might be involved in both the initiation and progression of cartilage lesions.

In contrast with this idea, some results involving animal experiments showed that the subchondral plate and trabeculae underneath cartilage were thin when cartilage loss started; only in later OA development, the bony sclerosis occurred, along with thickening of both areas of bone [40,92].

These discordant findings may be explained, first by the fact that Radin et al. immobilized the joint during the experiments. This mechanism may not represent the usual process of OA development. Furthermore, without the proper means of monitoring the disease evolution, it may be impossible to tell whether cartilage or bone lesion comes first. Also, cartilage loss and bone sclerosis could be two independent processes, both consequences of increased mechanical stress [53].

Although controversial, these findings underlined the importance of understanding the pathophysiologic sequences and consequences of OA pathology and contributed to the gradual shift on the characterization of OA from a cartilage centred view towards a whole joint disease.

1.3 Biomarkers in Osteoarthritis

Biomarkers are key elements in the evaluation of pathogenic processes, used as indicators of abnormal biologic processes and pharmacologic interventions efficacy.

Biomarkers of bone and cartilage quality may allow diagnosis in the early stages of OA. One example of a cartilage biochemical marker is the urinary levels of collagen type II C-telopeptide fragments (CTX-II) [123] that can be used as an indicator of cartilage degradation. Bone quality and bone strength have typically been measured by bone mineral density (BMD) or by bone histomorphometry measures such as trabecular thickness and trabecular number [109]. BMD obtained from dual energy X-ray absorptiometry scans is often used in osteoporosis research to measure bone quantity and it is not necessarily related to other aspects of trabecular bone, such as structure and quality.

For several decades, clinical research has relied on radiographic biomarkers for the diagnosis of OA. The traditional marker (US Food and Drug Ad-

ministration (FDA) approved) is the joint space narrowing (JSN) measured from radiographs. Besides JSN, the Kellgren & Lawrence (KL) [81] grading method is commonly used to define OA. The KL score measures the joint space width — and thus indirectly cartilage degradation — together with other OA features such as osteophyte formation and sclerosis [105]. Table 1.1 describes the KL grading system.

However, scores like JSN and KL are based on radiographs, where mainly bone and hard tissues are visible. The development of unbiased and more sensitive indicators of OA could lead to smaller clinical studies and make the development of drugs more efficient. Alternatively, more recent imaging techniques appear promising in measuring the morphologic and molecular state of cartilage [61] and bone structure [87, 100]. Some of these measures are commented in Section 1.5.

Table 1.1: The table describes the KL grading system that grades from radiographs the severity of OA [81].

KL 0	healthy	Normal
KL 1	early OA	Doubtful narrowing of joint space and possible osteophyte lipping
KL 2	OA	Definite osteophyte and possible narrowing of joint space
KL 3	OA	Moderate multiple osteophytes, definite narrowing of joint space, some sclerosis and possible deformity of bone contour
KL 4	advanced OA	Large osteophytes, marked narrowing of joint space, severe sclerosis and definite deformity of bone contour

1.4 Machine learning and texture analysis in imaging biomarkers

Improvement in imaging biomarkers requires the application of prior knowledge and a discriminative representation of the data. Because of large variations and complexity, it is hard to formulate analytic solutions for these tasks. Alternatively, solutions based on machine learning has brought significant

advances to numerous fields in medical imaging [133].

The goal of machine learning is to solve a given problem by applying knowledge acquired from an automatic learning process using a set of example data or past experience [2]. Some applications of machine learning include computer-aided diagnosis, image-guided therapy, image database retrieval, image segmentation, registration and annotation [133].

Texture analysis is an image analysis technique that quantifies the variations in intensity patterns, including variations that are imperceptible to the human visual system [80]. Texture can be analyzed in terms of model-based signal processing, variations of intensity, or structural elements. One example of a structural analysis is the study from Sørensen et al. [132], where a general texture-based machine-learning framework was used for measuring emphysema in CT images of the lungs. Another example is the work by Raundahl et al., where a feature set based on eigenvalues of Hessian matrix was used to quantify specific biological effects in the breast tissue [122].

Support for a texture analysis approach also comes from the investigation in [143] that combined a set of texture features to evaluate different tissues and their changes in multiple sclerosis studies. Their outcome endorses that a general set texture analysis can discriminate between different tissues and contribute to early diagnosis.

Until now, most of the texture features used in the analysis of trabecular bone have been based on spatial image statistics [6]. For example, the investigation presented in [139] employed four automated methods of texture analysis (grey level histogram, co-occurrence, runlength and gradient matrices) for structural characterization of trabecular bone.

Fractal analysis is another approach of spatial statistical analysis that has been computed to characterize variations in the structural network of the trabeculae bone. By applying fractal signature analysis (FSA) in digitised macroradiographs to quantify changes in the subchondral tibial trabecular bone in knee OA [19], Buckland-Wright et al. suggested that increased horizontal trabecular thickness occurred early and preceded the later changes in the vertical structures in the diseased compartment. The concept of fractal lacunarity analysis was introduced to provide parameters sensitive to bone micro architecture changes in aging. The authors proposed a lacunarity function to represent the variation of mass density of pixels in the image,

which was used to represent a standard for the evaluation of trabecular bone architecture.

1.5 Magnetic resonance imaging markers

Quantitative and semiquantitative measurements in magnetic resonance imaging (MRI) scans might allow improvement of a subjective evaluation. Quantitative measurements consider one or several parameters to objectively characterize the pathology, typically using a continuous scale. On the other hand, semiquantitative measurements typically include a combination of subjective and objective evaluations involving quantitative measures and visual assessment, often using an ordinal scale. A thorough review of quantitative and semiquantitative measurement methods in OA was presented by Eckstein et al. [45].

Current semiquantitative measures of the whole knee joint cover various structures within the joint. An example of a semiquantitative method is the whole-organ MR imaging score (WORMS), which has been used in several clinical trials and epidemiologic studies [111, 125]. Other whole-joint semiquantitative methods are the Knee OA Scoring System (KOSS) [86] and the Boston-Leeds OA Knee Score (BLOKS) [69]. All methods analyze OA features such as cartilage defects, bone marrow lesions, subchondral cysts, meniscal abnormalities and osteophytes. The features are scored by a trained radiologist, resulting in an overall score of the knee. Neither of the covered structures analyze the trabecular network of plates and rods directly.

Standard techniques of stereology combined with texture analysis of high resolution MRI have been used to quantify trabecular structure and to derive measures such as the mean intercept length, apparent trabecular bone area fraction, apparent trabecular spacing, apparent trabecular number and apparent trabecular spacing [87, 100]. These parameters are called "apparent" due to the limited spatial in vivo resolution of MRI. The study of Lin et al. [96] applied a thresholding technique based on regional intensity histograms to calculate these measures from images obtained from a 1.5 Tesla MRI scan. Their results demonstrate that MRI may be a very useful tool for bone structure and microarchitecture assessment.

Examples of MRI cartilage markers include: volume, thickness, smooth-

ness, curvature and homogeneity. The volume marker describes the quantity of cartilage, which might be normalized to the joint size [57,99]. Thickness is measured as the mean thickness of the cartilage sheet [32,129]. The smoothness relates to the fine-scale surface curvature and the curvature marker measures the global bending of the cartilage sheet [56,67]. The homogeneity measures the uniformity of the cartilage and can be computed as 1 minus the entropy, where the entropy quantifies the intensity histogram of the cartilage [114].

Chapter 2

PhD dissertation: Purpose and overview

2.1 Purpose

The main purpose of this PhD research was to investigate potential bone markers based on quantification of the tibial trabecular bone structure, in order to support early diagnosis and prognosis of knee OA. More sensitive and automatic markers can allow clinical studies with a smaller population or a shorter duration, both potentially reducing the cost of the study and supporting the drug development process. The studied methodology was based on universal machine learning and texture analysis of MRI scans.

Universal approaches are general methods designed for generic tasks, as opposed to tailor-made methods designed to solve specific tasks. Typically, these methods include a training phase where task-specific parameters are learned. Examples of medical imaging applications, using this approach can be seen in [132, 143].

2.2 Choice of region of interest

This project focused on the analysis of tibial trabecular bone structure (see Figure 2.1). Chapter 4 presents the quantification of the subchondral medial trabecular tibial bone. The region of interest (ROI) was defined as a region
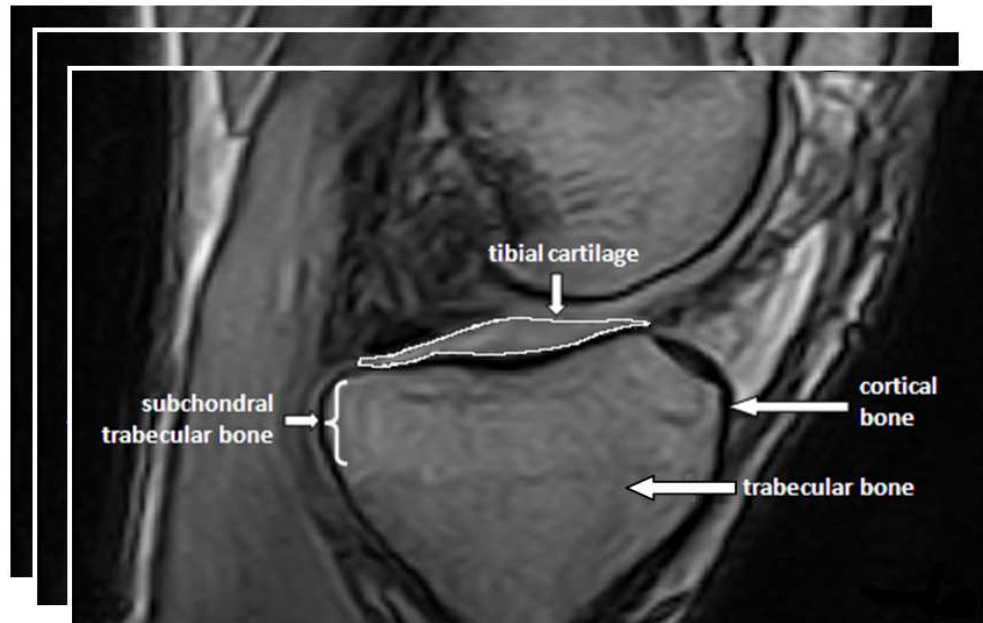
Figure 2.1: Sagittal slice of a 3D scan of a knee highlighting the trabecular bone, the subchondral trabecular bone, the cortical bone and manual outlines of the tibial cartilage. The subchondral bone is the region below the tibial cartilage, providing support for the cartilage of the articular surface. The cortical bone is the hard outer layer of the bone, while the trabecular bone is the porous, sponge-like inner structure.

right below the medial tibial cartilage with a small margin to the cortical bone, in an area expected to have relatively homogeneous biomechanical stress. Other chapters consider the whole tibial trabecular bone as the ROI.

The mechanism by which bone contributes to OA is until now poorly understood. Although it remains unclear how bone is involved in OA, as bone adapts to loads by remodelling to meet its mechanical demands, bone alterations likely play a role in OA development. Previous measurements have shown that reduced tissue hardness and alterations in microstructure of the trabecular bone tissue are associated with OA [30]. Furthermore, some speculation has stated for many years that bone may be a primary organ triggering OA [53].

By focusing on trabecular bone quantification, we intended to support the development of markers targeting the early structural changes in OA, which could boost dedicated research into treatments focusing early-stage

OA, where it may be reversible [8].

2.3   Choice of modality

Joint bone structure has previously been investigated mainly from radiographs, computed tomography (CT), micro-computed tomography (μCT), or high-resolution MRI's. However, μCT is only applicable *ex vivo* and radiographs are two-dimensional, generating projection artifacts. So, CT and mainly MRI are used increasingly in OA research.

Furthermore, radiographs are not very perceptive to change and early pathologic features of OA [70]. For example, the study of Jones et al. [77] indicated that a considerable amount of cartilage (1-13%) can be lost before radiographic OA (ROA) may be detected by JSN.

Alternatively, MRI enables 3D visualization and allows imaging of most tissues in the joint, including the trabecular bone structure and its changes [105]. When compared to radiographs, a recent MRI study [62] reported signs of early OA in the majority of knees in subjects without signs of ROA, suggesting that, compared to radiographs, MRI may be superior at detecting and hence understanding early OA of the knee in humans.

In this project we used low-field MRI scans of knees from an existing longitudinal study. Analyzing trabecular bone in a MRI with low resolution is a non-trivial task. Low-field MRI suffers from partial volume effects, hence tiny structures like the trabeculae are smaller than the voxel size and, therefore, jointly contribute to the voxel gray-scale value.

To deal with partial volume effects, statistical methods can calculate a local intensity threshold value that could distinguish voxels containing pure marrow from those partially occupied by bone [136]. Contrarily, we extract a generic set of features to capture the differential geometry structure of these intensities, potentially supporting the analysis of the intricate network of the trabecular bone as a whole. Previous work based on texture analysis has shown that characteristics of the trabecular architecture may be extracted without requiring rigorous segmentation between the individual trabeculae [75].

2.3.1 Quality of Images in low field MRI

Clinical MRI scanners have magnetic field strengths ranging from 0.1 to 3.0 Tesla. When the field strength increases, the signal-to-noise ratio improves and the potential resolution increases. In some high-field MRI scanner, the trabecular structure of the bone can be more visible. Therefore, it may be possible to directly monitor the trabeculae and determine its importance in the development and progression of OA [105]. However, these MRI scanners are costly and the scanning time increases with the resolution.

The lower resolution in a Low-field MRI implies that tiny structures like the individual trabeculae are not directly visible. The trabecular thickness range from 100 to 150 μm [83]; therefore, they are smaller than the voxel size, contributing jointly to the voxel gray-scale value. On the other hand, they are less expensive to purchase and operate, which make them attractive for both clinical studies and the general practitioner.

In terms of image quality, high-field MRI have shown better spatial and contrast resolution, but lower field strength adds fewer artifacts and smaller chemical shift artifacts [134]. Furthermore, low-field scanners are found in smaller versions than the whole-body high field scanner, designed to scan specific body parts such as the knee. This kind of scanner places the magnetic coil closer to the structure to be imaged, which increases the field homogeneity, as compared to the whole-body scanner [106].

Despite the differences, prospective studies [47, 84, 85] have shown that low-field MRI can achieve diagnostic accuracy comparable to high field MRI for detection of bone erosions, synovitis, medial meniscus, along with anterior and posterior cruciate ligament tears. The only exception was presented by Kinnune et al. [84], where the sensitivity for lateral meniscus lesions was considerably lower than in hight-field MRI.

2.4 Main contributions

The main contributions of this dissertation are:

- Investigation of aggregate markers based on measurements targeting different anatomical structures, the medial tibial cartilage and the subchondral tibial bone (Chapter 4).

- Comparison between different classifiers and different linear feature selection methods in texture analysis (Chapter 4 and Chapter 5).

- The improvement of a previous developed texture analysis framework [103] by the implementation of a new feature selection strategy (Chapter 5).

- A bone quantification method that potentially can foster the development of a new marker able to predict disease progression and identify patients most likely to progress (Chapter 6).

- Identification of the inferior part of the tibial bone as the most relevant region for prognosis of cartilage loss (Chapter 7).

- Preliminary findings suggesting a slight relation of the developed bone quantification marker to vertical trabecularization (Chapter 6 and 7).

2.5   Overview of the Thesis

After the two first introductory chapters, the main content of this thesis is presented in six chapters. Chapter 3 describes the data sets used in all experiments presented in this thesis. It includes the MR image acquisition, a short description of the study population and a description of the features extracted from the images. Chapter 4 presents the first experiments for diagnosis of OA, evaluating the subchondral medial trabecular tibial bone and using a previous version of the framework [103]. The new framework is presented in Chapter 5. This chapter also details our investigation in linear feature selection methods. Chapter 6 presents the clinical outcome of the new texture analysis framework. We analysed the whole trabecular tibial bone to quantify a prognosis bone marker potentially able to predict disease progression and identify patients most likely to progress. In Chapter 7, we describe a methodology to identify the most informative region of a ROI and show which region of the tibial bone is potentially more related to cartilage loss. Chapter 8 describes a preliminary investigation on sparse approaches for linear discriminant analysis and partial least squares applied for texture analysis. The last chapter, Chapter 9, concludes the dissertation by summarizing the research and providing a general discussion of its content.

Chapter 3

Image Data Aquisition

3.1 Study population

For the experiments described in this thesis, we used data sets consisting of MRI scans of both left and right knees from 159 test subjects in a longitudinal, community-based, non-treatment study [33]. After exclusion of scans due to acquisition artefacts, 313 knee scans remained in the diagnosis data set. The prognosis data set consisted of 268 scans, on account of the subjects that dropped out prior to follow-up.

All scans have been scored using the Kellgren & Lawrence score [81] determined from radiographs by an experienced radiologist. The score ranges from 0 to 4, where KL 0 indicates a healthy knee, KL 1 borderline ROA and KL 2–4 defines a knee with moderate to severe ROA. Figure 3.1 shows examples of textures from scans with different KL indexes and Table 3.1 shows the distribution of KL scores in the population. The table shows the healthy/ROA groups separated by a horizontal line, the percentage of knees in each one is 81% and 19%, respectively.

The population characteristics were: age 56 ± 16 (mean and standard deviation), BMI 26 ± 4 and 47% female. The subjects signed informed consent forms. The study was conducted in accordance with the Helsinki Declaration II and European Guidelines for Good Clinical Practice [60]. The study protocol was approved by the local ethical committee.
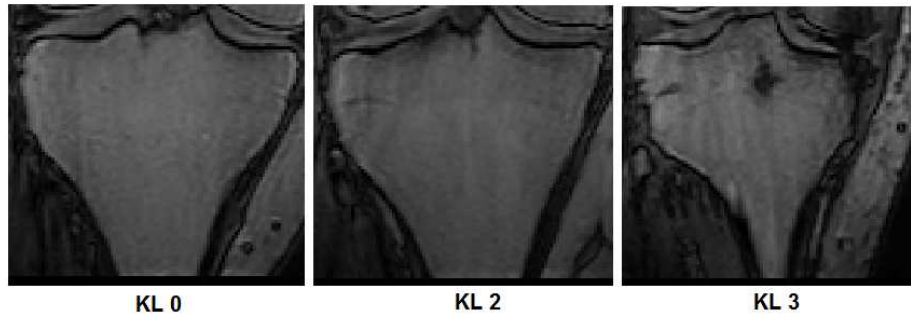
Figure 3.1: a) The figure shows examples of textures from scans with different KL indexes

Table 3.1: The distribution of the population: number of knees for each KL score. The horizontal line separates the healthy/borderline group (KL 0 and 1) from the ROA group (KL 2 to 4).

KL score	Num knees
0	158
1	94
2	31
3	29
4	1
Total	313

KL, Kellgren and Lawrence.

3.2 Magnetic resonance image acquisition and cartilage volume assessment

The scans were obtained in 2004 and 2006 at the Center for Clinical and Basic Research in Ballerup, Denmark. The MRI scanner was an Esaote C-Span low-field 0.18T extremity scanner. The key scanner parameters were Turbo 3D T1 sequence, 40° flip angle, 50 ms repetition time and 16 ms echo time. During scanning, test subjects were in a supine position with no load-bearing. Acquisition time was approximately 10 minutes.

The size of the scans was $110 \times 256 \times 256$ voxels and after automatically

removing boundaries that contained no information, the scan size was $110 \times 170 \times 170$ voxels. The spatial in-plane resolutions of the sagittal scans were 0.70×0.70 mm, with slice thickness ranging between 0.70 and 0.94 mm depending on joint size, where the most common distance was 0.78 mm making the voxels nearly isotropic. An example of a knee scan sagittal slice is visualized in Figure 2.1.

In order to assess the cartilage volume change over time, MRI was performed at baseline and 21 months later. The medial tibial cartilage volume was estimated by a fully automatic, computer-based process of segmentation [57] using the features described in Section 3.3.

3.3   Feature set

Texture Features: A total of 178 generic texture features were extracted from the images. The process intended to capture the differential geometry structure of the local intensities, which could allow the analysis of the network of the trabecular bone anatomy.

A generic feature set that has been demonstrated to provide good results for many patterns is the *N*-jet [55, 57, 121, 132]. The *N*-jet applies Gaussian derivative filters bank equivalent to the partial derivatives of a local Taylor series approximation up to order *N*. The partial derivatives calculated at a given image point and a given scale were used as a basis for representing different visual textures of the image. We included the 3-jet, based on the Gaussian derivative kernels including derivatives up to the third order.

Furthermore, to allow modelling of complex texture, gradient vector and magnitude and non-linear combinations of the Gaussian derivative features were included. Specifically, these combinations were the structure tensor [138] and Hessian eigenvalues and eigenvectors.

To provide basis for capturing anatomy and pathology of varying sizes, the texture analysis framework considered three different scales: 1, 2 and 4 mm. We also did experiments with scales 0.5, 2 and 8 mm and the results were qualitatively identical. A small scale was included to encompass, for example, the trabecular structure of the bone and larger-scale features were needed to handle larger structures such as BML that can be 5-20 mm in size [29], roughly corresponding to the support of a Gaussian filter at scale 4

mm. A detailed list of the extracted features is shown in Table 3.2.

Table 3.2: Feature set with: number of dimensions for each feature (2nd column), number of features extracted considering the dimensions and order of the derivatives (3rd column), the scales calculated (4th column) and the total of each feature considering also the scales computed (5th column).

Feature	Dim.	Num Features	Scales	Total
Intensity	1	1	-	1
Derivative (0th, 1st, 2nd and 3rd order)	3	20	1 2 4	60
Position	3	3	-	3
Position normalized	3	3	-	3
Structure tensor eigenvalues	3	9	1 2 4	27
Structure tensor eigenvectors	3	9	1 2 4	27
Hessian negative ridge	1	1	1 2 4	3
Hessian positive ridge	1	1	1 2 4	3
Hessian eigenvalues	3	3	1 2 4	9
Hessian eigenvectors	3	9	1 2 4	27
Gradient magnitude	1	1	1 2 4	3
Gradient vector	3	3	1 2 4	9
Third order derivatives	1	1	1 2 4	3
Total of features				178

Feature Scores: The features were calculated in each voxel. However, to capture both the feature level and variation across a ROI, we summarized each extracted feature in three possible scores: the mean, the standard deviation and the Shannon entropy. By including entropy, also aspects of non-Gaussian distributions could possibly be captured.

When extracting the features at three scales and calculating the three feature scores for each ROI, the total number of features was 534. This large, generic multi-scale feature bank contained linear and non-linear features including features invariant to rotation and scaling. Thereby, we hoped to allow quantification of significant bone structures visible in the images.

Each feature was normalized to zero mean and unit variance on the basis of the distribution of the training set.

Quantification of the subcondral medial tibial trabecular bone for diagnosis of OA

Even though the pathogenesis of OA is a complex mix of events in the whole joint, until recently the main focus has been on the articular cartilage. It is known that cartilage works closely together with the subchondral bone, also including the trabecular bone (see Figure 2.1). Therefore, analyzing them in combination looks promising [13, 79]. In this chapter, we present a framework for automatically analysing the structure of the subcondral medial tibial trabecular bone in low-field MRI for diagnosis of OA. We also present an investigation of an aggregate marker based on measurements

targeting different anatomical structures, the medial tibial cartilage and the subchondral tibial bone.

In summary, the framework initially determines a ROI within the tibial bone based on a cartilage shape model that has good agreement with manual segmentations [32]. At later stages, we use machine learning techniques, at first for selecting the most disease-related features; and second, to automatically learn from examples how to analyze a ROI recognizing the trabecular bone of a knee with OA. The outcome of the framework is a single value for each ROI, which is defined as the bone structure marker. This value indicates the probability of having OA. The evaluation of the developed method is presented in Section 4.2, where we compare and aggregate the developed marker with existing OA biomarkers from different modalities.

4.1   Bone structure analysis

4.1.1   Cartilage Shape Model

The goal is to automatically define the ROI (the subchondral trabecular bone), which includes the same anatomical region for all scans, maintaining anatomical correspondence. Since the structure to be covered is the subchondral trabecular bone, the ROI can be automatically extracted based on a cartilage shape model.

The segmentation of tibial and femoral cartilage was based on the method developed by Folkesson et al. [57]. The first step yields binary segmentations of the medial cartilages by a fully automatic voxel classification method. The voxel classifier performs an approximate nearest neighbor approach, not classifying all voxels, but focusing on the anatomical structure being analyzed. The algorithm starts from a set of randomly sampled voxels; it classifies them as either the object (medial cartilage) or background. If a voxel is classified as cartilage, it continues with classification of the neighboring voxels and this expansion process continues until no more neighboring cartilage voxels are found.

Second, the statistical shape model developed by Dam et al. [32] was fitted to the tibial segmentation in order to regularize the voxel classification results and to provide an anatomical coordinate system. The method used
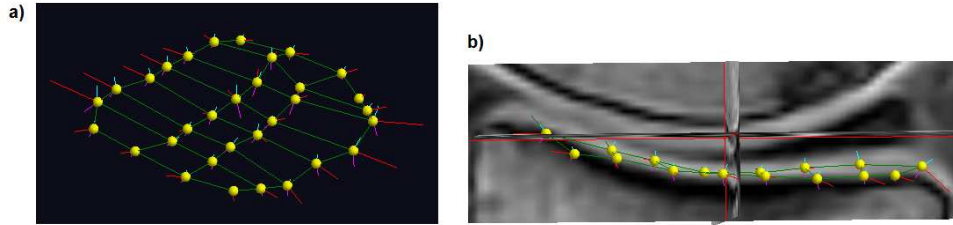
Figure 4.1: a) The m-rep shape model fitted to the segmented cartilage with the grid of medial atoms in yellow. b) Slices of a knee scan with the corresponding shape model.

an m-rep model to provide a smooth shape representation [112]. The trained mean model was initialized to match the binary tibial cartilage classification by translation to center of mass and scaling to the same volume. Then, the m-rep model was deformed to model the binary classification accurately while preserving the anatomical correspondence given by the shape model. The deformations were a combination of global similarity transformations, global modes of shape variation, along with local changes to the medial atoms and their attributes; all optimized by conjugated gradient descent. The global modes of variation were based on principal geodesic analysis [78]. The resulting cartilage shape model (see Figure 4.1) formed the support of the automatic bone ROI identification.

4.1.2 Extraction of the region of interest

The ROI was contained within the medial tibial condyle, in an area expected to have relatively homogeneous biomechanical stress. Thus, based on the cartilage shape model (presented in Section 4.1.1), the ROI was defined as a region right below the cartilage with a small margin to the cortical bone. The parameters were chosen by visual inspection of scans of knees with varying degrees of OA, ensuring that the ROI contained only trabecular structure.

We defined the ROI to cover the area from 5 mm to 15 mm below the cartilage sheet. This range was used for all scans. In the anterio-posterior direction, the region was defined from 15% to 95% of the cartilage sheet, and from 10% to 95% in the medio-lateral direction.

The shape model allows zero thickness within the cartilage sheet giving proper modelling of denuded area. Thus, the corresponding area of gaps in
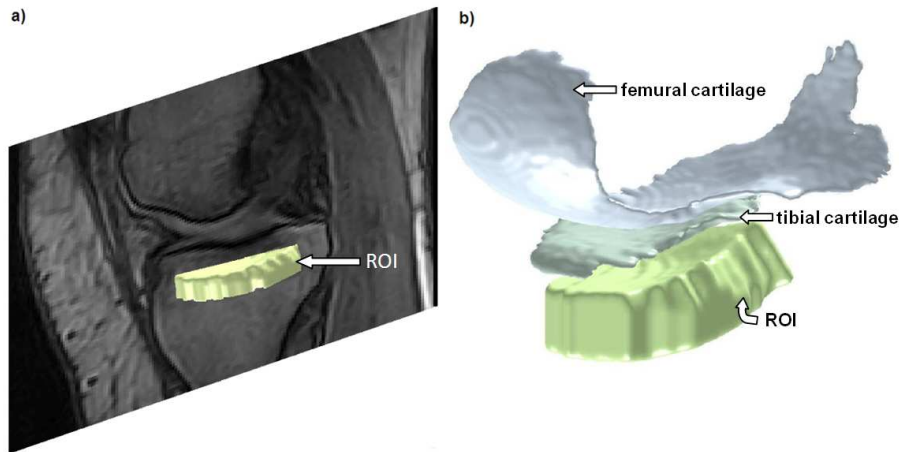
Figure 4.2: a) Sagittal view of a knee scan showing the tibio-femoral joint. The cortical bone appears as an almost black surface around the light trabecular bone within. The highlighted 3D area is the extracted ROI. b) Tibio-femoral segmentation showing each of the anatomical structures: the ROI extracted from the subchondral region of trabecular bone, the tibial cartilage and the femoral cartilage.

the cartilage sheet is included in the ROI. An example of an automatically extracted ROI is shown in Figure 4.2.

4.1.3 Classification and selection of features

Feature selection

Feature selection was performed to provide an appropriate feature set to the classifiers. The feature sets were selected by sequential floating forward selection (SFFS) [63,130]. SFFS is a classifier-dependent, suboptimal method, previously shown to have a performance comparable to the optimal methods [76,113]. The algorithm initiates with the empty feature set and iteratively includes the best additional feature in a greedy fashion. After each inclusion, SFFS excludes one or more features if the resulting feature set performs better than the previous feature set of the same size. We allow repeated inclusion of each feature, resulting in an increasing weight of the feature.

Potentially, the floating nature of the method can correct "mistakenly" added or removed features. Therefore, the nesting problems of other sub-

optimal feature selection methods, such as sequential forward selection, can be reduced [63].

To ensure a manageable computation time, a maximum number of features was set in the SFFS. When the maximum number of features was reached, the algorithm stopped. The overall best feature set was chosen as the set with the highest performance regardless of its size. Preliminary experiments showed that already at a feature set size of 10, the training area under the ROC curve (AUC) did not improve significantly. Therefore, the maximum number of features was set to 20, which makes room for the algorithm to float backward and forward before reaching the maximum number of features.

Classifier

Since the feature score distributions were unknown, six different supervised classifiers were evaluated: linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), nearest-neighbor (NN), k-nearest-neighbor (kNN), weighted nearest-neighbor (wNN) and weighted k-nearest-neighbor (wkNN). The performance was evaluated with respect to ability to classify the knee scans as healthy (KL 0–1) or osteoarthritic (KL 2–4).

Except for the weighted nearest-neighbor schemes, the classifiers are standard choices. The weighting of the k-Nearest-Neighbor scheme [23] can improve the generalization, especially for high-dimensional data sets with few samples as is the case here. The scheme simulates bootstrapping by weighting the sorted neighbors, such that the closer the neighbor, the higher the weight.

Since the data is unbalanced with respect to the number of knees for each class (healthy/OA), cost functions such as the classification accuracy would be inappropriate, whereas the AUC is suitable [131]. Therefore, the AUC was chosen as the performance measure.

Biomarker definition

For a given classifier, the SFFS algorithm results in a feature set. This classifier delivers a single output for each ROI — we define this as the bone structure marker.

4.2 Evaluation methodology

To handle the classical curse of dimensionality due to the limited number of scans and the high-dimensional feature bank, we perform the evaluation using cross-validation.

4.2.1 Cross-validation strategies

When evaluating the performance of a given feature set for a given classifier, two performance measures are considered: the training AUC and the generalization AUC. The training AUC describes how well the chosen features explain the training data, and the generalization AUC describes how well the found features separate new data [131].

The goal is to maximize both performance measures by choosing the features that entail a high training AUC when doing feature selection. A typical assumption is that a high training AUC will result in a high generalization AUC. However, the training AUC will normally be larger because of some degree of overfitting — particularly in a high-dimensionality feature space, low-sample size setting.

To avoid excessive overfitting, designing the feature selection so the chosen feature set generalizes well is important. Typically, the data is divided into three sets: training, validation and test [63]. The training set forms the training data for the classifier, and the validation set is used for testing the performance of a given feature subset, resulting in the training AUC. Finally, when the optimal feature subset is found, the generalization of the feature set is tested by classification of the test set.

This strategy is a typical example of cross-validation. To encompass the diversity of the data, particularly with limited data, the data are often randomly divided into the three above-mentioned sets N times, hence N evaluations are performed. Typically, the three sets are chosen of equal size - this specific case we will denote by CV3.

Another special case of cross-validation is leave-one-out (LOO), where a single sample from the data set is used as validation data and the remaining samples as training data. This is repeated such that each sample from the entire data set is used once as the validation data. In order to evaluate this approach, instead of considering the entire data for evaluation, we consid-

ered each sample from the test set in turn and the remaining data was defined as training set. The LOO scheme is expected to generalize better because the number of training samples is increased, but it is computationally more expensive than CV3.

4.2.2 Framework implementation

The overall framework for diagnosis of OA based on the trabecular bone structure of the tibia is summarized in the following algorithm:

1: **for all** knee scans **do**
2:    Calculate ROI
3:    Extract features
4: **end for**
5: **for** each evaluation **do**
6:    Divide data in training, validation and test sets
7:    Normalize features for each set
8:    **Do** SFFS until maximum features reached. The classifier is trained on the training data and evaluated on the validation data.
9:    **return** Training AUC for feature sets of size 1 to 20
10:    Choose the feature set with highest training AUC to compose the bone structure marker specification
11:    Calculate generalization AUC on test data
12: **end for**
13: Calculate median results for all evaluations

The feature selection experiments compared the performance of the six classifiers when doing feature selection: LDA, QDA, NN, kNN, wNN and wkNN.

For each classifier, both CV3 and LOO were done in order to validate the idea that LOO performs better than CV3 when limited data are available and the feature space is high-dimensional. For each, a total of 500 evaluations were done, where the data set was randomly divided into subsets.

For CV3, the data were divided into the three sets with 1/3 in each. Specifically, this resulted in 105 training scans, 104 validation scans and

104 test scans. In LOO, the data were divided into two sets, where 2/3 were used for training/validation and 1/3 for testing. LOO feature selection included 209 scans, of which 1 was iteratively chosen for validation and 208 for training. For both schemes, the overall training and test performance were calculated as the median of the 500 AUCs.

Parameters

Feature selection was performed with a maximum of 20 features. Determining k for the kNN classifier was done using the rule of thumb: $k = \sqrt{n}$, where n is the total number of training samples [131]. For CV3: $k = 10$, LOO: $k = 14$. Additionally, the ROI description was given in Section 4.1.2 and the feature scale parameters were given in Section 3.3.

4.2.3   Combination with other biomarkers

Previous research show that combining biochemical and imaging markers can result in good aggregate markers that improve diagnosis and prognosis of knee OA [34]. Therefore, in order to better encompass the complex pathogenesis of OA, the developed bone structure marker was evaluated in combination with other types of biomarkers related to OA diagnosis: a biochemical marker and cartilage MRI markers.

The biochemical marker was the urinary levels of collagen type II C-telopeptide fragments (CTX-II) [123]. Five different tibial MRI cartilage markers were included: volume, thickness, smoothness, curvature and homogeneity. All markers were automatically computed based on the cartilage segmentation presented in Section 4.1.1.

The combination of the mentioned OA biomarkers, except the bone structure marker introduced in this paper, was previously evaluated by Dam et al. [34].

The performance of the individual biomarkers was analyzed and compared to the developed bone structure marker. Furthermore, it was evaluated an aggregate marker that combines the developed bone marker with the mentioned OA biomarkers. The biomarkers were evaluated by the LDA LOO classification scheme described in the previous section.

Table 4.1: The median AUC of the CV3 and LOO feature selection evaluation for each classifier. In parenthesis is the standard deviation.

Classifier	Training AUC	Generalization
LDA CV3	1.00 (0.011)	0.79 (0.065)
LDA LOO	0.98 (0.011)	0.82 (0.051)
QDA CV3	1.00 (0.006)	0.76 (0.072)
QDA LOO	0.99 (0.006)	0.81 (0.058)
NN CV3	0.93 (0.033)	0.63 (0.054)
NN LOO	0.90 (0.028)	0.65 (0.059)
wNN CV3	1.00 (0.008)	0.75 (0.064)
wNN LOO	0.98 (0.013)	0.77 (0.065)
kNN k10 CV3	0.99 (0.014)	0.76 (0.061)
kNN k14 LOO	0.96 (0.016)	0.79 (0.060)
wkNN k10 CV3	1.00 (0.009)	0.77 (0.063)
wkNN k14 LOO	0.97 (0.014)	0.81 (0.060)

AUC, area under the ROC curve; CV3, cross-validation; LDA, linear discriminant analysis; LOO, leave-one-out; NN, nearest neighbor; QDA, quadratic discriminant analysis; wNN, weigthed nearest neighbor; kNN, k-nearest neighbor; wkNN, weigthed k-nearest neighbor.

4.3 Results

4.3.1 Classifiers evaluation

The training and generalization AUC for each classifier is shown in Table 4.1. Across all classifiers, the median training AUC varies from 0.90 to 1.00 and the generalization AUC from 0.63 to 0.82.

Comparing CV3 and LOO, the results indicate that LOO had worse training AUC but improves the generalization AUC. This means that the span between training and generalization AUC decreased as expected when increasing the training data.

When doing weighting of the nearest neighbors in the NN and *k*NN

classifiers, in general both training and generalization AUC increased — most pronounced for w*k*NN.

The classifiers w*k*NN, QDA and LDA has highest generalization AUCs. Despite the simplicity and linearity, LDA performs best among all classifiers with AUC 0.82. It is worth noting that LDA LOO had lowest standard deviation for generalization, although the differences are perhaps not statistically significant.

The training AUC in the CV3 evaluation for these three classifiers are all 1.0 and somewhat lower in the LOO evaluations whereas the generalization AUCs are higher in the LOO versions. This means that for the given size of the training set, robustness against overfitting is a central issue.

4.3.2   Aggregate biomarker

The alternative biomarkers targeting cartilage from both MRI and biochemistry were compared against the bone structure marker that was achieved using the LDA classifier (that performed best in the above evaluation). Additionally, LDA combinations of these markers were evaluated. The evaluation was performed using the LOO scheme.

The median generalization AUC for each biomarker and for the aggregate biomarkers are shown in Table 4.2. The statistical significance of AUC scores (columns 3 to 5) and differences in-between were tested using DeLong's test [41]. This nonparametric test approach takes into account the implicit correlation between the ROC curves by estimating a covariance matrix based on the generalized U-statistics.

The individual biomarkers spanned from AUC 0.58 for cartilage volume ($p = 0.416$) to AUC 0.79 for cartilage smoothness ($p = 0.001$). The biochemical marker CTX-II, the cartilage markers smoothness and curvature along with the developed bone structure marker, all had AUC scores above 0.70. However, the bone structure marker had the highest AUC among all the individual markers.

Combining all biomarkers except the developed bone structure marker resulted in AUC 0.82, which had the same result as the bone structure marker performed individually. It is worth noting that the aggregate biomarker also including the bone structure marker resulted in AUC 0.85.

Table 4.2: The median generalization AUC and the p-values for each biomarker. The p-value for the biomarker vs. the developed bone structure biomarker and last, the p-value for the biomarker vs. the aggregate marker. The horizontal line separates CTX-II and cartilage markers from our developed markers.

Biomarker	AUC	p-value	p-value vs. bone structure	p-value vs. aggregate all
CTX-II	0.71	0.006	< 0.001	< 0.001
Cartilage volume	0.58	0.416	0.008	< 0.001
Cartilage thickness	0.60	0.346	0.007	0.001
Cartilage smoothness	0.79	0.001	0.525	0.286
Cartilage curvature	0.75	0.009	0.365	0.118
Cartilage homogeneity	0.66	0.093	0.083	0.021
Aggregate cart+CTX-II	0.82	< 0.001	0.544	0.453
Bone structure	0.82	< 0.001	-	0.176
Aggregate all	0.85	< 0.001	0.176	-

AUC, area under the curve; CTX-II, C-telopeptide of type II collagen.

The table also shows the statistical significance of the AUC scores both individually (compared to chance) and in comparison to the bone structure marker and the all-including aggregate marker.

4.4 Discussion

The results demonstrated the feasibility of separating healthy and OA knees based on the trabecular bone structure. Trying several classifiers revealed that for this problem, a linear classifier performed slightly better than the quadratic classifier and the non-linear classifier *wk*NN. We operate with high-dimensional feature spaces, so when doing classification with a limited data set, the feature space will be sparsely populated. When doing classification, the simple, parametric LDA makes a general rule to separate data based on

all training samples. *k*NN, however, based the classification on only the *k* nearest neighbors, that may be distant from the test sample and possibly not sufficiently well describe the distribution locally. For this task, the simple, parameterized classifier apparently is more robust and generalizes better.

The results showed that each of the individual markers provided some OA diagnostic ability. The best individual cartilage marker, cartilage smoothness, had AUC of 0.79 and the biochemical marker (CTX-II) had AUC 0.71. In comparison, the bone structure marker reached AUC of 0.82, showing the potential of the developed marker for OA diagnosis. Furthermore, due to the fully automatic framework, the provided marker is well suited for large, longitudinal clinical studies.

However, the complexity and heterogeneity of OA make it unlikely that a single marker will allow a comprehensive quantification, suggesting the investigation of aggregate markers based on measurements targeting different anatomical structures. The results demonstrated that measurements from MRI and measurements from biochemical markers are complementary and allow superior aggregate markers for OA.

The aggregate marker including the bone structure marker was superior, albeit not statistically significantly, than the aggregate marker including only CTX-II and the cartilage markers for OA, developed by Dam et al. [34]. The increase in diagnostic ability is presumably due to the combination of the complementary aspects of bone and cartilage measurements.

Our study had some limitations. The parameters used to define the ROI were established visually from the data. To allow investigation of other regions within the bone, a possibility is to base the ROI on a full joint shape model as opposed to a cartilage model. This would better facilitate investigations into differences between medial and lateral condyles, between regions with more or less load-bearing, and between cartilage and subchondral bone regions. A further possibility is to evaluate the marker framework for discrimination at different levels of OA.

Much more research is needed to determine the clinical features that the bone structure marker describe. Potential candidates are textural patterns related to bone marrow lesions, growth zone characteristics and the architecture of the trabeculae even if the individual trabeculae are not visible.

In conclusion, we have demonstrated the possibility of automatically

quantifying a trabecular structure marker related strongly to the presence of OA. The final bone structure marker resulted in AUC 0.82 for diagnosis of radiographic knee OA, which was superior to existing OA cartilage biomarkers.

Chapter 5

Linear feature selection and framework improvement

The experiments described in this thesis uses a generic bank of 534 features (see Chapter 3 for more details). The drawback of a bank of features is a potentially high-dimensional representation of data. Beyer et al. showed that increasing the number of features, and thus the dimensionality of the data, leads to a loss of meaning for each feature and possibly decreases the model accuracy [10].

In many situations, a large number of features are strongly correlated and do not introduce new information to improve the ability to analyze the images. Feature selection and feature extraction (as feature space transformation) are effective approaches for dimensionality reduction (DR). Feature selection is the process of reducing the dimensionality by selecting a subset of the original variables. Feature extraction is a transformation of the input

data which possibly reduces the dimensionality by some functional mapping of D-dimensional data into a d-dimensional data, where ($d \leq D$) [97], e.g., by principal component analysis (PCA). These techniques can decrease the model complexity and eliminate noisy subspaces.

Partial least squares (PLS) regression is another example of multivariate data analysis technique that has been used for dimensionality reduction [68, 82, 119]. In this chapter, we introduce a robust PLS-based dimensionality reduction method in a texture analysis framework applied to diagnosis of knee OA. The framework was applied to classify between healthy subjects and OA patients by quantification of the tibial knee bone structure. The framework implemented the following steps: Initially, a generic bank of texture features was extracted from the images. As a pre-processing step to the dimensionality reduction, the outliers were identified and eliminated. Next, a feature relevance index (FRI) was applied to the original features to order them according to their importance to the model. Then, the method iteratively selected the features and transformed them to an orthogonal space. In the final step, the transformed features were used as input to a linear classifier that returned for each knee the probability of having OA.

5.1 Background

5.1.1 Linear Discriminant Analysis for Classification

Fisher linear discriminant analysis (LDA, [89]) is a classical classification tool that considers both classification and dimensionality reduction. The method decomposes the total covariance into between-class covariance and within-class covariance and maximizes the ratio of between-class to within-class scatter. The low dimensional discriminative space is estimated based on the resulted linear transformation.

This method is known to perform quite well in low-dimensional settings. However, in high dimensions, LDA cannot be applied directly because the covariance matrix estimate is singular and cannot be inverted. Although an approximated inversion can be calculated, in this case, the LDA method can suffer from high variance, resulting in poor performance [137].

To deal with this, the framework presented in this paper used PLS to apply

two dimensionality reduction approaches to the data set: feature selection and feature extraction. The resulted reduced space was sent as input to the LDA classifier. The following sections briefly explain the PLS method and describe some PLS based feature selection approaches.

5.1.2   Partial Least Squares

Multivariate data analysis techniques like PCA and PLS have a common principle; they are constrained to find a linear transformation that produces mutually uncorrelated components.

PCA is a classic technique used for dimensionality reduction by feature extraction. Concerning classification problems, we defined the data matrix as $\mathbf{X}$ (referred to as predictive or explanatory variables in factor analysis) and the classification classes as vector $\mathbf{Y}$ (the dependent or response variables). PCA linearly transforms the original features into uncorrelated features. These new features are ranked, loosely speaking, according to the amount of variability of $\vec{X}$ they explain. In this approach, no importance is given to how each feature may be related to the classes.

PLS does not require matrix inversion to obtain the coefficient matrix. Rather, the factors are computed by successive 1-D linear regression, assuming that the relationship between features and the classes is influenced by a few underlying variables, called latent variables or factors. The features and classes are assumed to be realizations of these underlying variables [142]. Therefore, PLS regression suggests the use of supervised dimensionality reduction by considering both $\mathbf{X}$ and $\mathbf{Y}$ information. This regression returns a linear combination of the features, the latent factors ($\mathbf{X}$-scores), which are used to predict the classes based on $\mathbf{Y}$-scores.

Figure 5.2 exemplifies how observations from a synthetic data set (see the data scatter plot in Figure 5.1) are represented in terms of the principal component coefficients created by PCA and the latent factors estimated by PLS. The PLS plot shows that PLS considered the correlation of the variables to the classes in order to compute the latent factors. The first variable had a weak relation to the class separability, so when computing the first latent factor, PLS almost ignored this variable while PCA considered the three variables nearly equally.
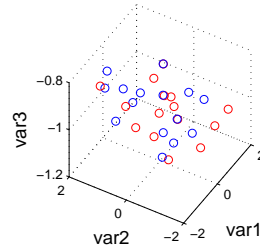
Figure 5.1: Synthetic data scatter plot: var1 is a variable with no correlation to the classes, var 2 is a random variable and var 3 is a variable with strong correlation to the classes.
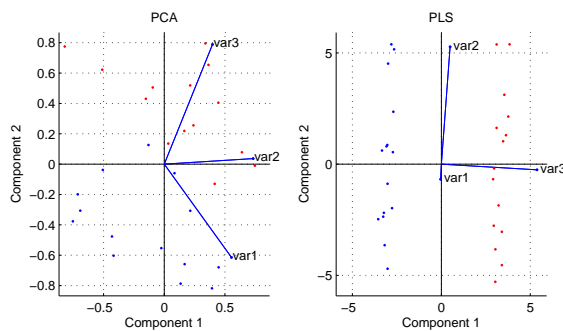
Figure 5.2: The first plot shows the axes representing the first two principal components extracted by PCA and how each observation is represented in terms of those components. The magnitude of the blue vectors represent each observed variable's contribution to the first component. The second plot shows how each observation is represented in terms of the latent factors extracted by PLS.

Barker et al. [7] presents a formal statistical explanation clarifying that the dimensionality reduction provided by PLS is determined by between-groups variability, while by PCA is determined by total variability. Therefore, PLS potentially performs better than PCA for dimensionality reduction when classification is the ultimate goal.

The PLS model also supports selection of the latent factors; considering h the total number of factors generated by the regression, by selecting only the first k ($k < h$) to compose the new feature set, one can reject the noisy information.

There are various algorithms for calculating PLS regression. The NIPALS algorithm [1] is the standard, but the SIMPLS [39] is very popular because

it is faster. Our framework used the SIMPLS algorithm for the PLS regression. In the following paragraphs, we briefly introduce the PLS regression for calculating the coefficients **B** and the factors **T**. Although PLS supports multiple response variables, we considered a single response **Y**.

For n samples and p predictors or features, a PLS regression model decomposes **X** and **Y** to produce the following bilinear representation of the data:

$$\mathbf{X} = \mathbf{T}\mathbf{P}' + \mathbf{E} \tag{5.1}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}' + \mathbf{F} \tag{5.2}$$

where the input data and classes are in $\mathbf{X} \in \mathbb{R}^{n \times p}$ and $\mathbf{Y} \in \mathbb{R}^{n \times 1}$ matrix, respectively. The **X**-scores, $\mathbf{T} \in \mathbb{R}^{n \times h}$, contain the transformed features in the orthogonal space while the matrix $\mathbf{U} \in \mathbb{R}^{n \times h}$ has the transformed classes (**Y**-scores). The scores matrices contain the information on how the samples relate to each other. $\mathbf{P} \in \mathbb{R}^{p \times h}$ and $\mathbf{Q} \in \mathbb{R}^{1 \times h}$ are the loading matrices, which represent the regression coefficient of **X** on each column of **U** and the regression coefficient of **Y** on each column of **T**, respectively. Loadings contain the information on how the features relate to each latent factor. The matrices **E** $\in \mathbb{R}^{n \times p}$ and $\mathbf{F} \in \mathbb{R}^{n \times 1}$ contain the residuals. The latent factors (equation 5.3) and regression coefficients (equation 5.4) are computed based on a weight matrix $\mathbf{W} \in \mathbb{R}^{p \times h}$ that expresses the correlation of each **X**-column with the **Y** variable. Thereby, entries in **W** with values close to zero express less important features.

$$\mathbf{T} = \mathbf{X}\mathbf{W} \tag{5.3}$$

$$\mathbf{B} = \mathbf{W}\mathbf{Q}' \tag{5.4}$$

PLS is typically not robust towards outliers due to use of least squares regressions. The regression can focus on the atypical observations instead of describing the model represented by the majority of the data [38]. Furthermore, the SIMPLS algorithm is based on the empirical covariance matrix, where outliers can have a damaging effect on the estimates.

To reduce the effect of outliers, some approaches, such as the method presented in [15], propose to replace the empirical covariance matrix by a

robust covariance estimator. In [37], Daszykowski et al. presented a robust version of PLS by applying a weighting scheme to down-weight the negative influence of outliers upon the model. The outliers were detected based on standardized leverage and residual distances exceeding a cut-off value. For a more complete review on the leading robust PLS algorithms, see the work by Kruger et al. [88].

An alternative approach to robust PLS algorithms is to detect the outliers and eliminate them before defining the final model. According to Wold et al. [141], the initial outcome of a PLS regression can be used itself to detect outliers. Moderate outliers can be identified by the residuals of **Y** and **X** (**E** and **F** in Equations 5.1 and 5.2). The guidelines from the authors claimed that samples that deviate outside of 4 times the standard deviation (SD) of the **Y**-residuals can be considered an outlier. For the **X**-residuals, one needs to summarize all *k*-values for each sample. This is proportional to the distance between the data point and the model plane in **X**-space, the so called DModX (distance to the model in **X**-space). A sample with DModX larger than 2.5 times the SD of the **X**-residuals indicates an outlier.

The strong outliers are found by analyzing the Hotelling T^2 of the latent factors, the scores **T** in Equation 5.3. The Hotelling T^2 is proportional to the leverage, which is a measure of the influence of a sample on the PLS model [49]. Samples with confidence level less than a certain percentage can be considered outliers. Usually the limit level is between 95% and 99%.

5.1.3 Feature Selection Based on PLS

When a model includes a high-dimensional set of features, usually several of them are correlated. Besides providing nearly the same information to predict the classes, highly correlated features can imply model convergence problems, overfitting and the "curse of dimensionality" in general [95]. Therefore, an important part of learning-based processes is to identify a subset of features weakly correlated to the classes. Several PLS-based feature selection approaches have been proposed [3, 104], mainly to estimate the features more related to the underlying latent data structure.

In a recent work, Li et al. [94] proposed the use of the absolute values of the regression coefficients of the PLS model as an index for evaluating the

importance of wavelengths of multi-component spectral data. The proposed algorithm, called competitive adaptive reweighed sampling (CARS) first ranks the variables according to their absolute PLS coefficients. Based on this ranking, the algorithm sequentially selects N subsets of variables from N Monte Carlo sampling runs. In each sampling run, a fixed ratio (e.g., 90 %) of samples is randomly selected as training data. Next, based on the regression coefficients, a two-step procedure is adopted to select the key variables. The first step uses an exponentially decreasing function and the second step uses an adaptive reweighed sampling method to remove the unimportant variables. Finally, a cross-validation (CV) approach is applied to choose the subset with the lowest root mean square error.

Another example is the interactive variable selection (IVS) [50] that modifies the PLS algorithm by doing a dimension-wise selective reweighting of single values in each column of the weight matrix **W**. The investigation presented two techniques that use a threshold, defined by CV, to replace some values in **W** to zero. Their experiments showed that the elimination of either small or large values in **W** improved the model, but no clear explanation was given justifying the elimination of large **W**-values. Moreover, the author used a small simulated data set and the method was in part interactively evaluated.

Wold at al. [141] introduced the variable importance in the projection (VIP), which is a score that summarizes the importance of each variable for the latent factors projections. The score is a weighted sum of squares of the weights in **W**, with the weights calculated from the amount of **Y**-variance of each PLS latent factor (see equation 5.5):

$$VIP_j = \sqrt{p \sum_{k=1}^{h} \left(SS\left(q_k t_k\right) \left(w_{jk} / \|w_k\|\right)^2 \right) / \sum_{k=1}^{h} \left(SS\left(q_k t_k\right) \right)} \qquad (5.5)$$

where $SS(q_k t_k) = q_k^2 t_k' t_k$.

Building on Wold's work, Bryan et al. [18] presented the MetaFIND application that implements a post-feature selection method based on VIP and correlation analysis of metabolomics data. The features were ranked, but the threshold that defines the selected features was a user-defined parameter.

The "greater than one rule" is generally used as a criterion for variable

selection, since the average of squared VIP scores is equal to 1, but Chong et al. [25] showed that the proper cut-off value may be higher than 1 under uneven class distributions, high correlation, or an equal coefficients structure. Their investigation also explored the nature of the VIP method compared with other methods.

Although PLS combined with VIP scores is often used when multi-collinearity is present among variables [50, 141], there are few guidelines about how to use it [25]. To address this issue, our developed framework introduces a robust PLS-based strategy for dimensionality reduction (DR) that includes outlier detection, feature selection and feature extraction.

5.2   Framework

The overall texture analysis framework is summarized in the following steps:

1. Segmentation of the region-of-interest
2. Features computation
3. Dimensionality reduction
    3.1 Pre-processing steps:
       a) Auto-scaling and initial PLS regression
       b) Initial feature ranking
       c) Identification of the outliers
       d) Re-computation of FRI
    3.2 Iterative forward feature selection
       a) Incremental feature selection
       b) Intermediate evaluation
4. Classification
5. Evaluation

The features computation (step 2) is described in Section 3.3. The steps 1, 4 and 5 are described in Section 5.3, where the application of the texture analysis framework to OA diagnosis is detailed. The following section (Section 5.2.1) explains the step 3, the proposed dimensionality reduction method.

5.2.1   The dimensionality reduction method

The DR method implemented four pre-processing steps and an iterative forward feature selection (FFS). The outcome of the pre-processing steps was a robust ordered set of features. Next, the iterative forward selection defined the number of transformed features (latent factors) to use in the final model and which features from the original set to use in the calculation of these latent factors.

Pre-processing steps

Auto-scaling and initial PLS regression: As the first step, the feature set was normalized to zero mean and a standard deviation of one. Then, an initial PLS regression was calculated considering the given samples and features. The three next pre-processing steps were based on the output of this regression.

Initial feature ranking: The feature relevance index, *FRI*, illustrated in Figure 5.3 and defined in Equation 5.6, was computed:

$$FRI_j = \sum_{k=1}^{h} (|w_{jk}|)v_k \tag{5.6}$$

In Equation 5.6, w are the elements of the weight matrix $\mathbf{W} \in \mathbb{R}^{p \times h}$ and $|w_{jk}|$ returns the absolute value of this weights; v are the values in $\mathbf{V} \in \mathbb{R}^{h \times 1}$, which are the percentage of variance explained by each latent factor on both $\mathbf{X}$ and $\mathbf{Y}$ spaces, considering h as the number of PLS latent factors with the best evaluation.

The percentage of variance was computed based on the Pearson's correlation coefficient [124] between the sum of the features and the scores ($\mathbf{X}$-space) and between the classes and scores ($\mathbf{Y}$-space). After calculation, the index was normalized to length one.

The Equation 5.6 differs slightly from related work. Lindgren et al. [50] suggest weighting the $\mathbf{W}$-elements with their correlation to $\mathbf{Y}$ only.

Identification of the outliers: The guidelines from [141] were implemented to identify and remove outliers during the training phase: samples that
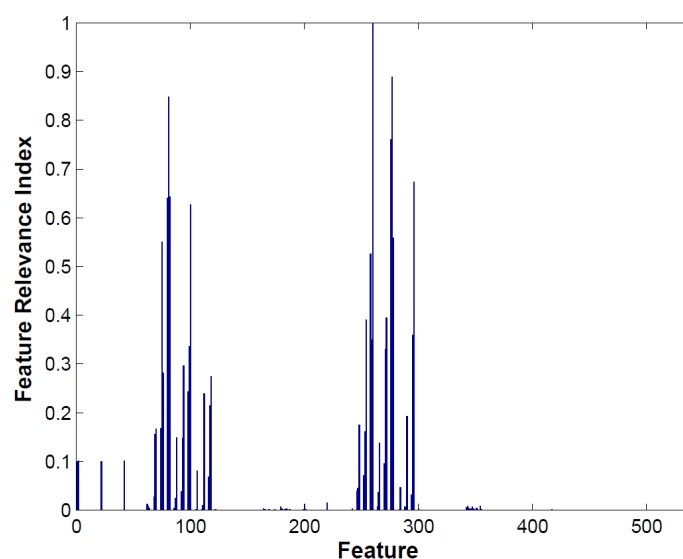
Figure 5.3: Bar graph of a FRI generated during a training phase. In this example, the original feature set has 534 features. The graph shows that the number of features identified as relevant was considerably less than the initial ones

deviate more than 4 times the SD of the **Y**-residuals and more than 2.5 times the SD of the **X**-residuals and with confidence level less than 0.99 on the Hotelling T^2 distribution of the **T** scores.

Re-computation of the FRI: To get a more reliable rank, this step re-computed the FRI. Basically, a new PLS regression was computed using the data without the outliers and considering only the most important features selected in the initial PLS regression step (only the first half of relevant features from the ordered index) and the outcome was used to re-compute the index.

Iterative FFS and optimization of the number of latent components

Finally, using a forward selection approach, this step generated a new feature set and defined the number of latent factors to be used in the final model. Starting from an empty set, the algorithm sequentially added the original features to the set, one at a time, considering the order established by the FRI. In each iteration, the algorithm used the current feature set to re-compute

the latent factors and an intermediate evaluation defined the number of latent factors. To avoid overfitting, the intermediate evaluation used cross-validation. The combination of selected features and number of latent factors with best evaluation were used in the final model.

5.3   Experiments

In this section, we specify the application of the developed framework to OA diagnosis, introducing the ROI definition and the applied classification and evaluation methods.

5.3.1   ROI Definition

A voxel classification algorithm, designed for cartilage segmentation [57], was generalized to also segment the tibial knee bone. From the segmented binary (mask-like) image, we applied a morphological erosion of approximately 2 mm to remove the outer layer of the bone corresponding to the cortical bone. The remaining was the trabecular bone, which was defined as the ROI (Figure 5.4).

5.3.2   Classification and Evaluation

To empirically evaluate the feature selection strategy, the data set was partitioned according to Monte Carlo CV [20] and to account for uneven class sample sizes, the samples were selected under the restriction that the number of OA and healthy subjects was proportionally distributed across the two sets. Apart from this, the distribution was random.

The DR method used leave-one-out CV on the training set to define the number of PLS latent factors and the number of features to be used on the final model.

For classification, the framework used Fisher LDA (described in Section 5.1.1) and for evaluation, we measured the area under the ROC curve (AUC). Since the data was unbalanced with respect to the number of knees for each class (healthy/OA), cost functions such as the classification accuracy were inappropriate.

Figure 5.4: Automatically segmented tibia bone in light gray and the region-of-interest (the trabecular bone) in gold.

5.4   Evaluated Methods

For comparison purposes, we evaluated 7 different methods. All the methods had as input the 534 original features generated in accordance with Section 3.3. With exception of experiments using support vector machines (SVMs, [28]), the output of all the other methods was considered as input to the Fisher LDA classifier.

To investigate dimensionality reduction by feature extraction methods, we first evaluated the performance of PCA and PLS without an additional feature selection step. A training phase using cross-validation defined the number of principal components and the number of latent factors for PCA and PLS, respectively.

Next, we evaluated the performance of the proposed PLS-FFS method (Section 5.2.1). The training phase defined the selected feature set and the number of latent factors to be used in the final model.

As an alternative approach, we added a stopping criterion to the FFS step

considering only the features with FRI greater than 10^{-4}, which corresponds to exclude all features with index value roughly equal to zero.

We also experimented replacing our feature selection step by the CARS algorithm [94] to compare the performances.

As a baseline for comparison, we considered two linear SVMs methods. First, we applied soft margin SVMs [28] to the full feature set. The regularization parameter, typically denoted by C, was determined for each of the 100 Monte Carlo CV folds independently by a grid-search procedure on logarithmic scale. Stratified 10-fold CV, partitioning the training subsample from the 100-fold CV, was employed to evaluate the different regularization parameters.

Second, we considered linear SVMs combined with recursive feature elimination (RFE, [64]). Recursive feature elimination is a common iterative feature selection technique for SVMs. In each iteration, a SVM is trained. In the resulting linear classifier there is one coefficient (or weight) per feature. The features are ranked according to the corresponding weights squared. Then the features with the lowest ranks are discarded. We employed a popular variant called SQRT-RFE, in which the number of features that are removed equals the square root of the number of features used in the previous SVM training [58]. Because there is no reason to assume that the regularization parameter is independent of the (number of) features, we adjusted C using a grid-search relying on 10-fold CV as described above in every iteration of the RFE procedure. The feature set yielding the lowest average test error in the 10-fold CV is regarded as the solution of the RFE process. For all SVM experiments, we used the Shark library [74].

5.5   Results

The results reported in Table 5.1 show the potential of feature space reduction of the evaluated methods. All the evaluated methods had p-values less than 0.0001, which indicates the AUC's (second column) were significant larger than an AUC of 0.5.

By efficiently reducing the number of original features, one can reduce the general model complexity. The third column shows the median of original features actually requested by the final model. The forth column shows

Table 5.1: Evaluations of the different methods. The last line shows the evaluation of the method with the additional stopping criterion: feature relevance index (FRI) greater than 10^{-4}. The third and fourth columns show, per each CV evaluation, the median of original features used and the classifier input dimensionality. The fifth and the last columns show the number of features used across all evaluations and the correspondent percentage of the original feature set.

Method	AUC	features per CV	DR per CV	features all CVs	%
PCA	0.89	534	60	534	(100%)
PLS	0.90	534	8	534	(100%)
SVMs	0.90	534	-	534	(100%)
SVM-RFE	0.90	54	-	450	(84%)
PLS-CARS	0.87	48	22	534	(100%)
PLS-FFS	0.92	102	26	194	(36%)
PLS-FFS FRI$>10^{-4}$	0.92	91	26	182	(34%)

the input dimensionality of the LDA classifier. One can argue that since LDA performs better in low-dimensional settings, by reducing its input dimensionality the overall classification performance can be improved.

To cope with the limited number of samples, the data was evaluated using CV. The drawback of the CV approach is that for N iterations, N estimations of the parameters will be generated, being unlikely to produce a single set of selected features that best discriminate the samples. We evaluated the stability of the feature selection algorithms by analysing how different training sets generated from the same distribution affect the relevance of the features. The fifth and the last columns show the total number of features selected across all CV iterations and the correspondent percentage of the original feature set.

Considering the total number of features selected across all CV iterations, the proposed DR method used 36% of the original feature set; the alternative approach, with stopping criterion, selected 34%; whereas the other approaches used considerably more. These results demonstrated that our feature selection method decreased the model complexity, which can potentially contribute to a better understanding of the anatomical characteristics of data being analysed.

The experiments applying only feature extraction (PCA and PLS) had similar generalization AUC, around 0.90. By including the feature selection step in the PLS method, we could identify the subset of features actually used by the texture analysis framework without decreasing the model accuracy. The proposed method, PLS-FFS, reached an AUC of 0.92. Although the performance was only slightly better, the results suggested that our method detected some features that hinder the classification model instead of providing useful information.

In a comparison, the feature selection step with the CARS algorithm selected in median a relativity small number of features (9% of the features on each CV iteration), but it decreased the accuracy of the model to AUC 0.87. The outcome of a small feature subset seems to be a characteristic of the method, called the key variables. However, the detection of OA based on texture analysis apparently needs a larger set of informative features. Furthermore, considering the number of the features across all CV iterations, the algorithm used all of the original features. In fact, the selected features in each CV iteration had a high variability, which indicates instability of the method.

The linear SVM applied to all features showed a good performance with an AUC of 0.90. This did not change when using in median only 10% of the features after applying RFE. While the RFE-SVM training and feature selection was much faster compared to the PLS-based approaches, the accuracies of PLS-FFS were not reached and the feature selection was less stable in the sense that only 16% of the features were discarded in all of the 100 CV folds.

5.6 Discussion

Applied to diagnosis of OA, our texture analysis framework reached a diagnostic ability of AUC 0.92 by analyzing the tibial bone structure. For comparison, a recent study analyzing a linear combination of several morphometric and structural cartilage markers in the same population scored AUC of 0.84 [34]. Although the studies analyzed different anatomical structures, the results showed that the presented method captured the texture changes and had diagnostic ability superior to other biomarkers of OA.

A drawback of our method, due to the iterative strategy, was the rela-

tively high time consumption for evaluating the selected feature set when it reached more than 100 features. In order to overcome it, we defined an additional stopping criterion evaluating only features with FRI greater than 10^{-4}. The diagnostic ability was similar to the original proposal and, as a direct consequence of the anticipated stopping, the total number of selected features across the CV iterations were only 34% of the original feature set, compared to 36% without this stopping criterion. Although the addition of a stopping criterion was effective in reducing the computational time, the minimum relevance index value may need to be reviewed for different data sets.

5.7 Conclusion

We presented a robust dimensionality reduction method based on PLS regression that combined outlier detection, feature selection and feature extraction. The results illustrated that our method performed effectively analysing texture separating healthy subjects and OA patients: the diagnostic ability of our method reached AUC of 0.92 compared to 0.84 in a previous study. Furthermore, we obtained a considerable reduction in the number number of features used by the final model (in median 19% of the original set).

Future improvements to the developed method include evaluating the framework with some robust PLS algorithm instead of excluding the outliers in a pre-processing step and investigating learning algorithms with a sparsity constraint, in particular linear classifiers with L1 or L0 norm regularization. Furthermore, validation on another independent study population is key to verify the results.

Besides this, the clinical challenge that remains for future research is an examination on the relationship between the selected feature sets and the pathological features they can represent, e.g., BMLs, osteophytes, bone erosions, subchondral cysts, bone attrition, among others.

Chapter 6

Quantification of tibial trabecular bone for prediction of tibial cartilage loss

Research on MRI analysis has provided methods for quantification of the multiple components of the knee joint, besides revealing significant structural associations of sub-regions of the subchondral bone area with cartilage loss [16, 117, 118]. For example, the work presented in [72, 140] analysed MRI scans to assess BML and to measure the change between baseline and follow-up in knee cartilage volume. Their results suggested that enlarging BMLs are strongly associated with more cartilage loss.

Additional investigations in these associations could lead to MRI markers of ROA progression, which may have advantages over existing methods.

Coordinated interdisciplinary work to design sensitive biomarkers able to identify patients most likely to progress would enable more rapid assessment of new treatments and structure-modifying therapies in clinical trials [71].

In this chapter we used the longitudinal study described in Chapter 3 to investigate the quantification of OA and prediction of medial tibial cartilage loss by analysis of the tibia trabecular bone from baseline MRI scans of knees. The framework described in Chapter 5 was used to quantify two bone markers, one for diagnosis of ROA (KL>1) and another one for prognosis of rapid cartilage loss.

This chapter also presents a preliminary radiological reading of the high and low risk knees. This analysis put forward a hypothesis of which pathologies the bone marker could be capturing to define the prognosis of cartilage loss.

6.1   Brief review of the framework

To quantify the trabecular bone structure, initially a voxel classification method segmented the tibia bone. From the segmentation, we applied morphological erosion of approximately 2 mm to exclude most of the cortical bone. The remaining was basically the trabecular bone, which was defined as the ROI. The data set contained a generic set of 534 features, including Gaussian derivatives at multiple scales, extracted from the ROI. At next step, a machine learning approach was applied to deal with two different tasks involving two classes each. Firstly, to discriminate between healthy and OA patients and afterwards to segregate the slow and rapid progressors of cartilage loss. To individually deal with the tasks, the framework applied PLS feature extraction followed by forward feature selection to identify the linear combination of the features that best discriminate the two classes. The selected features were employed to train a LDA classifier. Basing on a cross-validation strategy, the outcome of the classification was a single value for each knee, representing at first, the classified probability of having OA, and afterwards the classified probability of rapid cartilage volume loss. For the first task, the outcome of the classification was defined as the diagnosis bone marker and for the second one, the prognosis bone marker. The overall texture analysis framework is described in Section 5.2.

6.2   Analysis of the relation between the bone marker and pathological features

According to the developed prognosis bone marker, 30 baseline knees were selected for further analysis: 15 belonging to the high risk group of cartilage loss and 15 belonging to the lowest risk group. An experienced radiologist (Harry Genant) reviewed and scored the radiographs of these knees. The radiograph analysis was blinded, so the specialist did not know how much a specific knee lost in cartilage or whether it was considered high or low risk by the bone marker. The specialist focused particularly on the medial joint region: medial joint space narrowing (MJSN), medial osteophyte (MPHYT), osteopenia (OP) and vertical trabecularization (VTRAB).

The measurements considered a 8-scale (0=normal, 0.5=normal worse, 1=mild, 1.5=mild worse, 2=moderate, 2.5=moderate worse, 3=severe and 3.5=severe worse) based on the Altman and Genant scoring method [4]. During the analysis, the scores were grouped in only two categories due the small number of samples: one category with values less than 1 and another with values equal or more than 1.

6.3   Statistical Analysis

As an initial validation of the proposed methodology, we apply the texture analysis methodology to OA diagnosis of the baseline images. The first experiment evaluated the diagnostic ability by classifying the population in healthy (KL=0 or KL=1) and ROA (KL>1). The outcome was the probability of having ROA.

The second experiment investigated the feasibility of predicting the rapid/slow progressors of cartilage loss based on the texture analysis of the trabecular bone structure of a baseline knee image. In this case, the outcome of the classification was the probability of rapid cartilage volume loss. The progressors were designated by the median of cartilage loss calculated across the population. The ones below the median were defined as slow progressors and the rest as rapid progressors. The texture analysis framework used the area-under-the-ROC (AUC) to evaluate the classification. Since the data was

unbalanced with respect to the number of knees for each class (healthy/ROA), cost functions such as the classification accuracy was inappropriate to this dataset.

In order to determine the accuracy of the method encompassing the diversity of the data, we used cross-validation by randomly dividing the data into independent training and generalization sets 300 times. The model was generated using a training sample and the generalization set was used to evaluate the model for unseen samples. Hence 300 evaluations were performed, one for each generalization sample set. The final AUC was computed by the median of the generalization AUC's. This strategy estimated how well the model can perform on future independent data set. The potential presence of confounding factors was evaluated by multivariate regression. Based on the results, the bone structure marker was normalized using the residuals of this regression.

6.4   Results

The results of the confounding factors analysis are shown in Table 6.1, where the odds ratio (OR) and the 95% confidence interval (CI) represent the evaluation of the diagnostic and cartilage loss prediction ability of gender, BMI and age. The following subsections present the results for the main tasks of this paper: diagnosis of ROA, prognosis of cartilage loss and the relation between the bone marker and some pathological features.

Table 6.1: Confounding factors analysis for diagnosis and prognosis of cartilage loss.

Factor	Diagnosis		Prognosis	
	OR	CI	OR	CI
Age	1.88	1.2−2.9	0.64	0.4−1.0
Gender	1.15	0.7−1.8	1.72	1.1−2.8
BMI	4.04	2.5−6.5	1.67	1.0−2.7
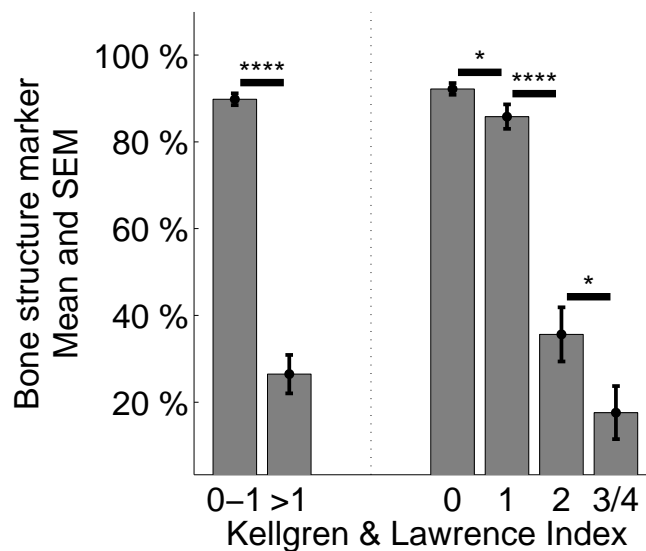
Figure 6.1: The probability of being healthy defined by the bone structure marker. The mean of measurements are shown (with bars illustrating the standard error of the mean, SEM) for the healthy and diseased groups, and, to the right of the dotted line, for each KL score. The levels of statistically significant separation were marked by stars: four stars correspond to $p < 0.0001$ and one star to $p < 0.05$.

Diagnosis

The ability to separate ROA (KL>1) and healthy knees of the bone structure marker reached a generalization AUC of 0.92 (p<0.0001) prior to normalization for the confounding factors and AUC 0.86 (p<0.0001) after correction. Figure 6.1 shows the cross-sectional separation.

Prognosis

By quantifying the bone structure, the framework predicted the rapid/slow progressors with OR of 3.9, with 95% CI 2.4 - 6.5. Age, gender and BMI did not affect the prognosis performance after correction. Figure 6.2 shows the actual cartilage volume loss separated out by median and tertiles of the bone marker. The medians of the subjects classified as slow and rapid progressors
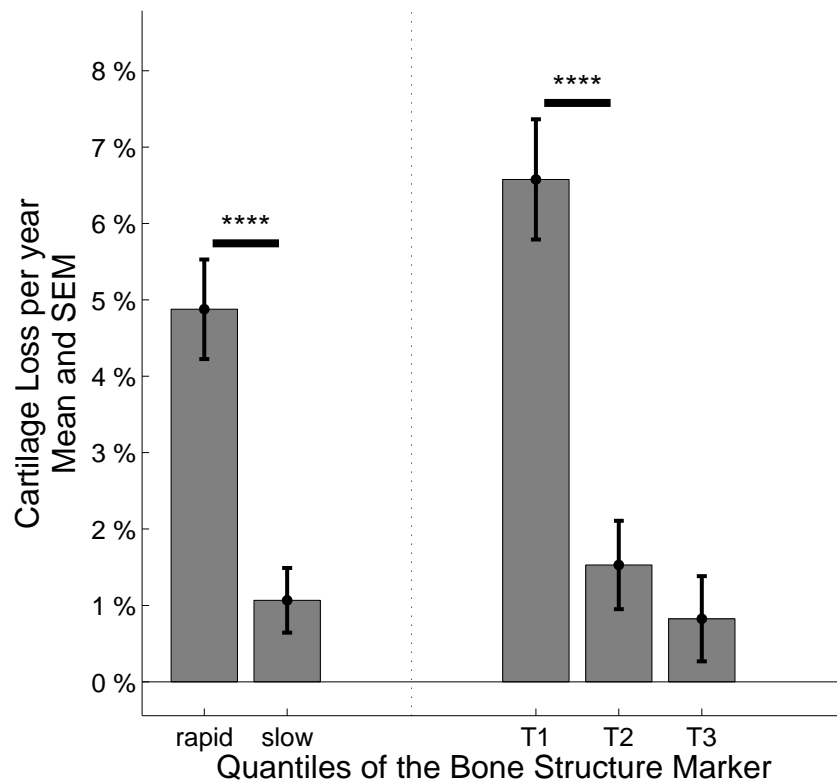
Figure 6.2: Cartilage volume losses separated out by above/below the median
of the bone marker on the left of the dotted line (4.9% and 1.1%) and tertiles
on the right side (6.6 %, 1.5% and 0.8%).

were 1.1% and 4.9% per year respectively.

We also evaluate the classification of progressors by tertiles. The bone
structure marker could segregate patients experiencing the greatest risk of
cartilage loss, 6.6% in the first tertile. The OR for the prediction of the first
tertile against the third one was OR 6.5 (CI $3.4 - 12.6$) prior to correction
and OR 5.6 (CI $3.1 - 11.2$) after normalization in relation to the confounding
factors.

Relation between the bone marker and pathological features

Table 6.2 shows odds ratios for the relations between the radiograph features evaluated and the cartilage loss and the relations between the features and the bone structure marker. Although not significant, the vertical trabecularization presented the highest relation to cartilage loss and the bone marker. As a remark, the segregation of the same samples using the bone structure marker had OR of 16.0 (CI $2.7 - 95.7$).

Table 6.2: Odds ratios and confidence intervals of the relations between the radiograph features and both the cartilage loss and the bone marker of 30 samples, 15 belonging to the top high risk group of cartilage loss and 15 belonging to the lowest risk group.

Measure	Cartilage loss		Bone marker	
	OR	CI	OR	CI
MJSN	1.00	$0.2-4.6$	1.00	$0.2-4.6$
OP	1.83	$0.7-9.3$	1.83	$0.4-9.3$
VTRAB	7.00	$1.1-46.0$	2.36	$0.6-12.4$
MPHYT	1.75	$0.4-8.0$	1.75	$0.4-8.0$

MJSN: medial joint space narrowing, OP: osteopenia, VTRAB: vertical trabecularization, MPHYT: medial osteophyte

6.5   Discussion

6.5.1   Related studies on prognosis of cartilage loss

So far, there is no clinically determined "gold standard" measure that quantifies trabecular bone for prognosis of cartilage loss. However, there is a substantial body of work on cartilage markers and their association with cartilage loss. For example, Hunter et al. [71] used longitudinal data from Boston Osteoarthritis of the Knee Study (BOKS) to determine whether biomarkers of cartilage turnover could serve as predictors of cartilage loss on MRI. They assessed the baseline levels of cartilage degradation and synthesis products by means of assays for type I and II cleavage by collagenases (Col2:3/4Cshort or C1,2C), type II cleavage only with Col2:3/4Clongmono (C2C), type II syn-

thesis (C-propeptide), the C-telopeptide of type II (Col2CTx), aggrecan 846 epitope, and cartilage oligomeric matrix protein (COMP). They concluded that with the exception of COMP, no other tested biochemical markers of cartilage synthesis and degradation facilitated prediction of cartilage loss. If changes in cartilage turnover in patients with symptomatic knee ROA were associated with cartilage loss, they did not appear to affect systemic biomarker levels.

In another study, Dam et al. [31] investigated whether the level of baseline urinary excretion of C-telopeptides of type II collagen (CTX-II) could predict progression of knee ROA as defined by radiographic signs and whether it could predict cartilage loss. Their results suggested that uCTX-II was a suitable prognostic marker, confirming previous reports claiming that CTX-II is associated with both the prevalence and the progression of ROA [123].

Studies that analyzed compartment-specific relation with cartilage loss, showed that subregions of the bone area is also a revelant structure for this investivagation. Eckstein et al. [46] investigated individual radiographic features (JSN, osteophytes, sclerosis, among others) and MRI cartilage morphometry features (specifically denuded subchondral bone area and cartilage thickness). The results showed that beside low cartilage thickness, knees with medial femoral subchondral bone sclerosis and medial denuded subchondral bone areas at baseline displayed significantly higher cartilage loss than those without. The canine study of Boileau et al. [14] also demonstrated that there was a highly significant correlation between total knee cartilage volume loss and subchondral bone hypersignal.

In another study, Roemer et al. [59] analysed if presence of baseline synovitis and effusion in knees without ROA predicted future tibio-femoral cartilage loss. The investigation used the longitudinal Multicenter Osteoarthritis Study (MOST). They concluded that baseline synovitis in knees without ROA does not predict cartilage loss, but predicts joint effusion instead. However, baseline effusion, which reflects synovial activation, predicts structural progression in subjects without ROA.

Pelletier et al. [110] investigated the correlation between knee cartilage volume loss and specific compartments such the menisci (tear and extrusion) and subregions of bone including subchondral bone, the tibial plateaus and femoral condyles. Their study showed that meniscal damage and bone

marrow changes are the features most closely associated with subregional cartilage loss. The greatest cartilage volume loss was found in the medial compartment, and risk factors included female gender, JSW, meniscal lesions and bone changes at baseline.

6.5.2 Our results

MRI markers of ROA progression represent an alternative for structural change which may have advantages over existing methods of measuring structure. Our investigation revealed that disease-related bone structure characteristics were reflected in the low-field MRI appearance. Even though the individual trabeculae were not visible, the outcome of our investigation demonstrated the possibility of analyzing the network of the trabecular bone as a whole and from this automatically to quantify a bone structure marker strongly associated with the presence of ROA.

Our first experiment validated the texture analysis framework in terms of ability to diagnose healthy and ROA subject. For this analysis, we consider the KL score as the "gold standard", where KL greater than 1 indicated ROA and healthy otherwise. The outcome showed that the applied analysis reached a diagnostic ability of AUC of 0.92. In a study using the same population [36], the joint space width had AUC of 0.73 while the best individual marker, cartilage roughness, had AUC of 0.80. A linear combination of several morphometric and structural cartilage markers, cartilage longevity, scored AUC 0.84. The results indicated that our framework had diagnostic ability superior to other biomarkers of ROA.

The second experiment evaluated the proposed bone marker applied for prediction of cartilage loss. Our results of the prognosis analysis showed that the tibial trabecular bone MRI texture can segregate ROA patients experiencing the greatest risk of cartilage loss (first tertile) from the ones with least loss (see Figure 6.2). As suggested by Pelletier et al. [110], patients from the first tertile are of particular interest from a clinical perspective as they are likely to have the worst prognosis and are, therefore, at greater risk of surgical intervention for joint replacement.

The development and validation of prognosis markers may accelerate the pace of therapeutic development. The advantages of this approach are

the use of low-field MRI equipments, which is cheaper to install, maintain and operate when compared to high-field; and a fully automatic computer-based framework, making the developed marker well suitable for large, longitudinal clinical studies.

6.5.3 Relation between the bone marker and pathological features

We presented an initial analysis of the relation between the bone structure marker and some pathologic radiograph features. With a small number of samples, it was not possible to achieve a high confidence level, which restricted the analysis of pathological findings. Nevertheless, the slight prominence of the vertical trabecularization in the results suggests further investigation on this relation. Despite the non-conclusive results, the experiment put forward a hypothesis of which pathologies the bone marker could be capturing to define the risk of cartilage loss.

6.5.4 Limitations

In this study, the relatively small number of subjects and mainly the high percentage of the population (80%) in early stages of ROA suggest that the findings need to be validated on other populations with more subjects in later stages of ROA. Specifically, the conclusions are more accurate for progression during the early stages of ROA. Furthermore, we suggest validation of our findings on high-field MRI scans, which may allow a more accurate and precise cartilage volume measurement and tibial trabecular bone quantification. Also, further research is required to integrate cartilage and bone into a comprehensive approach that can increase our understanding of the disease processes.

6.6 Conclusion

The applied texture analysis demonstrated the possibility to capture tissue changes and demonstrated the potential for automatically predicting the cartilage volume loss by analyzing the tibial trabecular bone structure. From a

theoretical point of view, the outcome underlined the importance of the bone to the understanding the disease processes. In practical terms, the developed quantification can contribute to the development of a new marker able to predict disease progression and identify patients most likely to progress. Our findings underlined the importance of the bone for diagnosis and prognosis of OA. It is aligned with the gradual shift on the characterization of OA from a cartilage centered view towards the whole joint system approach.

Chapter 7

The most informative texture region of the tibia for predicting cartilage loss

This chapter is based on the following manuscript. The introduction was rewritten to avoid redundancy when comparing to Chapter 1. The image data acquisition and feature computation were omitted since they were detailed in Chapter 3. Apart from these changes, the contents are similar.

J. Marques, D. M. J. Tax, M. Loog, E. Dam, "The most informative texture region of the tibia for predicting cartilage loss - A multiple-instance learning approach", Journal of IEEE Transaction on Medical Imaging (submitted).

Some of the interesting contributions of MRI to OA investigations are methods for quantification of the multiple components of the knee joint, which has enriched the analysis of the joint structure changes over time. Some approaches [117, 118] have revealed significant structural associations of subregions of the subchondral bone area with cartilage loss.

Additional support for this analysis comes from Chiba et al. [24]. The authors analysed structural features of subchondral trabecular bone in MRI scans of knees with OA in order to study OA progression. As cartilage

area decreased in the medial joint, some structural features increased in the medial tibia (bone volume fraction and trabecular thickness), while others decreased in the lateral tibia and femur (bone volume fraction, trabecular thickness and connectivity).

In view of these results, there are indications that anatomical features related to cartilage loss affect different subregions of the bone. One could argue that relevant insights can arise by analysing these regions separately. The work of Roemer et al. [126], for instance, analysed separately 14 subregions of the knee joint in order to score BML size and cartilage status in the same subregions. The subregions and scores were assessed manually. They found that the absence of BMLs was associated with a decreased risk of adjacent cartilage loss while subregions with new and progressive BMLs exhibited a high risk of cartilage loss at follow-up.

Motived by this interpretation, the work presented in this chapter analysed baseline images of knee MRI to explore subregions of the trabecular tibial bone, identifying the ones more related to cartilage loss and to classify subjects in rapid or slow progressors.

We analysed the texture of the ROI, the trabecular tibial bone, to quantify the structural elements related to cartilage loss. Traditionally, image quantification is based on the whole ROI at once [102, 133]. Contrarily, we quantified subregions of ROI separately, hopefully capturing the different pathological features occurring in each region of the bone and for this task we used multiple-instance learning (MIL).

Generally, MIL extends classical supervised classification in which every object is described by a single feature vector or instance to a description based on an arbitrary set of vectors, also called a bag. One of the first MIL algorithms was published by Dietterich et al. [42]. The idea was to find an axis-parallel hyper-rectangle that contain at least one instance from each positive bag and simultaneously exclude all the instances from negative bags. The authors applied the proposed algorithms to a drug discovery problem, where molecules were classified by looking at their shape statistics.

In our case, each subregion was defined to be one instance and a bag held all instances over a full ROI. We employed strict MIL in which the final decision is governed by the state of the worst subregion. That is, if the learning algorithm identified some tissue change related to cartilage loss in

at least one region (instance) of the whole ROI, the subject was classified as rapid progressor of cartilage loss. Afterwards, we evaluated the frequency with which the regions were selected as most relevant. In order to apply the proposed strategy, the MIL method had to provide accurately the positive instances and allow one to find the concept related to the positive bags. Section 7.1 gives more details about MIL and describes the employed MIL approaches.

Multiple-instance learning in medical imaging

In general, medical imaging applications require use of prior knowledge and a good representation of the data. Supervised machine learning techniques learn models using labelled data [2]. Traditionally, a training data set, consisting of data and label pairs for each sample, is used to build a classifier that can predict output labels for unseen samples.

As an alternative, MIL techniques extend this setting to include problems where the samples (bags) contain a group of unlabelled patterns (instances). In this case, the training set contains unlabelled instances belonging to labelled samples.

Applications of MIL include computer-aided detection [98,120] and medical imaging solutions [12,43,44], among others. Though no explicit mention of the term MIL is made, [98] and [120] present basic solutions to computer-aided detection tasks from chest radiography and mammography, respectively, that can be formulated as MIL problems. The work presented in [43] defined roughly segmented ROIs of Breast ultrasound image as a bag and subregions of the ROI were considered as the instances of the bag. They applied MIL for classification of tumours into benign or malignant. Another method for abstracting the task to multi-instance representation is presented in [12]. The authors proposed a novel classification approach for automatically detecting pulmonary embolism from computed-tomography-angiography images and also provided an approach for the problem of learning with multiple positive instances.

Relationship between bone quantification and pathological features

Our previous study in the same population [102] (see Chapter 6) have suggested a slight prominence of the vertical trabecularization in subjects classified as rapid progressors of cartilage loss. Although not significant, the vertical trabecularization presented a relation to cartilage loss and to the previous defined bone marker. The OR for the relation to cartilage loss was 7.0 with the 95% CI between 1.1 and 46.0. The relation to the previous bone marker had OR 2.4 (CI 0.6 − 12.4).

In this chapter, we present an analysis of the same baseline radiological readings in order to investigate whether the bone marker would also show some relation to vertical trabecularization when defining the most relevant region for prognosis of cartilage loss.

7.1 MIL for image classification

Formally in multiple-instance learning, a training set consists of a set of bag labels $y_1, ..., y_m$, associated to bags $X_1, ..., X_m$, where each bag contains a set of n instances, $X_i = \{x_{i1}, ..., x_{in_i}\}$. Note that not all bags have the same n. Typically, the instance space $\chi \in \mathbb{R}^d$ is the d–dimensional Euclidean space, and $\mathbf{Y} \in \{-\mathbf{1}, \mathbf{1}\}$ corresponds to the labels negative and positive. The goal of classical MIL algorithms is to train a instance function $h(X) : \chi \longrightarrow \mathbf{Y}$ that will accurately predict labels $y_{ij} \in \{-\mathbf{1}, \mathbf{1}\}$ for novel instances x_{ij}.

Based on this instance classification, the elementary assumption is that a bag is positive if at least one of the instances in that bag is positive, formally: $y_i = \max_j\{y_{ij}\}$. The fundamental challenge in MIL is that one does not know which of the instances in a positive bag are actually positive and which ones are not.

The following subsections introduce the employed MIL methods. The major reason why we chose the mi-SVM [5] was because it allows one to find the concept that will label individual instances correctly. This property is required for our strategy of identifying the most relevant region of the tibia. Originally, one of the main points of MIL was to identify a concept that relates the positive bags to their instance attributes, however, not all of the recent

MIL methods provide this feature. For example, MILES (Multiple-Instance Learning via Embedded instance Selection) [22] considers more than one target concept and each one can be related to either positive bags or negative bags. Therefore, we opted to apply mi-SVM for identifying the most relevant region and to compare both approaches, mi-SVM and MILES, for the task of prediction of cartilage loss.

SVM for MIL

Support Vector Machine (SVM) [28] is a learning method that constructs a hyperplane supported by training data points, called support vectors, which can separate the data into distinct classes. The optimal hyperplane is the one with maximum margin separating the classes. Andrews et al. [5] generalized this concept and proposed the mi-SVM for MIL problems.

The mi-SVM explicitly treats the instance labels y_i belonging to positive bags as unknown variables. The goal then is to maximize the instance margin jointly over possible label assignments. The optimal discriminant function should ideally classify the bags according to the following rule: if a bag is negative ($Y_I = -1$), then all its instances are negative ($y_j = -1$ for all instances $x_{ij} \in X_i$). On the other hand, if a bag is positive ($Y_I = 1$), then at least one instance in the bag is positive. Hence, the positive half-space of the hyperplane should have at least one instance from every positive bag, while all instances belonging to negative bags should be in the negative half-space.

MILES

The above approach assumes that a bag is positive if and only if at least one of its instances is positive. The MILES [22] considers the situation that a negative bag may contain positive instances as well. Furthermore, it extends the idea from the Diverse Density proposed by Maron et al. [101] by assuming that there may exist more than one target concept and each one can be related to either positive bags or negative bags.

In order to determine the target concepts, MILES specifies a probability based on the similarity between a bag B_i and a potential concept x (see equation 7.1). The measure is defined by the instance x and the closest instance in the bag.

$$Pr(x|B_i) \propto s(x, B_i) = \max_j \exp\left(-\frac{\|x_{ij} - x\|^2}{\sigma^2}\right),\qquad(7.1)$$

where σ is a predefined scaling factor.

The set of selected target concepts C is defined simultaneously with the bag classification process. Considering all concepts $x^i \in C$, each bag in the training set can be mapped into an instance-based feature space $\mathbb{F}_c$ via the mapping $m(B_i) = [s(x^1, B_i), s(x^2, B_i), ..., s(x^n, B_i)]^T$, where each feature of a bag is its similarity with the concepts $x^i \in C$. Therefore, the coordinates of a given bag represent the bag's similarities to various instances in the feature space.

The dimensionality of the feature space can be high, since it is proportional to the number of training instances. To deal with it, the algorithm applies a feature selection step based on the 1-norm SVM [11]. After the feature selection step, all the instances are classified according to their contributions to the bag classification.

7.1.1 Radiograph analysis

In order to investigate the potential relationships between the outcome of the texture analysis framework and known pathological features related to OA, 30 baseline knees were selected for further analysis. Considering the median of cartilage loss of the population, 21 knees belonged to the high risk group of cartilage loss and 9 belonged to the lowest risk group.

An experienced radiologist (Harry Genant) reviewed and scored the radiographs of these knees. The radiograph analysis was blinded, so the specialist did not know how much a specific knee lost in cartilage or whether it was considered high or low risk by the bone marker. The specialist focused particularly on the tibia bone and medial joint region: medial joint space narrowing (MJSN), medial osteophyte (MPHYT), osteopenia (OP) and vertical trabecularization (VTRAB).

The measurements considered an 8-scale (0=normal, 0.5=normal worse, 1=mild, 1.5=mild worse, 2=moderate, 2.5=moderate worse, 3=severe and 3.5=severe worse) based on the Altman and Genant scoring method [4]. During the analysis, the scores were grouped in only two categories due the small number of samples: one category with values less than 1 and another
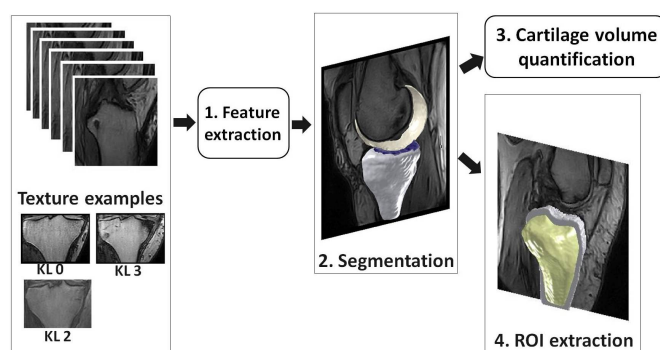
Figure 7.1: The initial framework steps: (1) feature extraction, (2) segmentation of the medial tibial cartilage and the tibia bone, (3) medial tibial cartilage volume quantification and in (4) automatically segmentation of tibia in light gray and the region-of-interest (the trabecular bone) in gold. The figure also shows examples of tibia in different stages of Osteoarthritis, from KL 0 (healthy) to KL 3 (severe diseased).

with values equal or more than 1.

7.2   MIL Framework for texture analysis

The initial framework steps included (Figure 7.1): features computation, segmentation of the cartilage and trabecular tibial bone, and cartilage quantification. Chapter 3 has more details about these steps. The ROI definition followed the same approach described in Section 5.3.1. The only difference refers to the morphological erosion. The experiments described in this chapter applied a erosion of approximately 5 mm to remove the cortical bone.

The remaining framework steps are described in the following sections. Subsection 7.2.1 defines the shape model and the associated coordinate system. Subsection 7.2.2 introduces the simultaneous evaluation and classification of the subregions. Finally, the strategy used to identify the most informative region of the ROI is explained in Section 7.2.3.
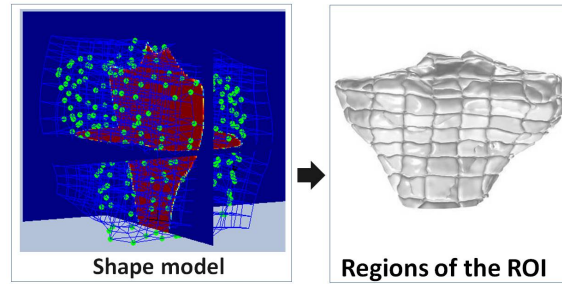
Figure 7.2: The figure shows an example of the shape model fitted to a segmentation of tibia. The coordinate system was defined based on a box of 15 : 10 : 8 in axial, sagittal and coronal direction respectively. The illustration shows a ROI divided in 400 regions.

7.2.1 Sub-regions feature set

Sub-regions partition In order to divide the ROI into sub-regions, a shape model was optimized to provide a coordinate system with approximate anatomical correspondence for the Tibia bone. The shape model was based on a point distribution model consisting of interior, exterior and boundary points. The correspondence was optimized in a framework inspired by previous work [35]. The shape model point grid defined a coordinate for each voxel of the ROI, allowing the ROI to be sub-divided (see figure 7.2).

Feature scoring In order to generate the final feature set for each region of the ROI, we summarized the extracted features by calculating the mean across each region. With this generic multi-scale feature bank, we aim to quantify significant differences in bone structures in the images.

7.2.2 Prognosis classification

The bags described the whole ROIs, the trabecular tibia bone, while each region of the ROI was defined as an instance of the MIL bag.

The label of the bags were defined according to the population median of the tibial cartilage volume change. The knees with cartilage loss above the

median were considered as the positive class (rapid progressors), otherwise they were assigned as negative class (slow progressors).

7.2.3 Identification of the most informative regions

Since mi-SVM allows to identify the positive instances of a bag and associates positive instances only with positive bags, we considered this approach for identifying the most informative regions. Therefore, after the classification, we counted how many times the mi-SVM selected each region as the witness to calculate the percentage of relevance of each ROI region.

7.3 Experiments

During the experiments, the MIL concepts were identified based on a randomly selected subsample of the data. The training set consisted of 168 samples. The remaining 100 samples was used as test set, for the model evaluation.

Due to the high variability of the data and the limited population size, the data set was randomly partitioned into training and test sets 100 times, hence 100 iterations of evaluation were performed. This sub-sampling technique is also known as Monte Carlo cross-validation [20].

The chosen performance measure for the prognosis analysis was the AUC. For the statistical analysis we included the OR with the respective 95% CI.

7.3.1 Definition of the most informative regions for prognosis of cartilage loss

A. Evaluation of prognosis of cartilage loss The first experiment evaluated the MILES and mi-SVM techniques in the binary classification problem of predicting the rapid/slow progressors of cartilage loss. The ROI was divided in 1, 3, 6, 8, 12, 24 and 45 regions. We considered that more than 45 regions could produce too small regions, which could break apart larger bone structures like BML. Besides, it could increase considerably the number of unlabelled instances, potentially leading to model convergence problems.

B. Evaluation of the most informative regions The second experiment analysed which regions were most related to cartilage loss. For this experiment, we considered the strategy described in the Section 7.2.3.

7.3.2   Bone marker statistics and pathological features

Based on the region most related to cartilage loss, we defined a bone texture marker for each knee. The classification was re-executed, now considering each bag with only two instances: one instance corresponded to the region more related to cartilage loss defined by the previous experiment and the other instance corresponded to the rest of the ROI. Based on the mi-SVM results, the bone marker for each knee was calculated as the median of posteriors, considering all the cross-validation evaluations.

Ideally, these experiments would be done in a different data set of images. But due to the small number of samples, we decided to evaluate it on the same population as a proof of concept that needs further validation.

C. Confounding factors analysis Using the bone marker, the potential presence of confounding factors (age, gender and BMI) was evaluated by multivariate regression.

D. Tertiles segregation   We also evaluate the ability of classifying progressors by tertiles, hoping to identify the patients with greatest risk of cartilage loss.

E. Relation between the texture analysis outcome and pathological features
After the definition of the bone marker, we investigated the relation between the bone marker and some pathological features related to OA.

7.4   Results

A. Evaluation of prognosis of cartilage loss   Table 7.1 shows the median AUC's and the standard deviation across the evaluations. The mi-SVM had predictive ability around AUC 0.70 when dividing the ROI in 3, 6, 8 or 12 regions. For 24 and 45 regions the AUC dropped to 0.63. The MILES

approach had a comparable performance evaluation only when analysing the whole ROI as the unique instance of the bag (first line). For the regions evaluation, its AUC was around 0.64.

Table 7.1: The median of AUC's and in parenthesis the standard deviation for 100 cross-validation evaluations.

Regions	MILES	mi-SVM
1	0.70 (0.01)	0.61 (0.17)
3	0.65 (0.05)	0.68 (0.02)
6	0.64 (0.07)	0.70 (0.01)
8	0.65 (0.06)	0.67 (0.03)
12	0.63 (0.11)	0.69 (0.01)
24	0.62 (0.15)	0.63 (0.09)
45	0.64 (0.07)	0.63 (0.10)

B. Evaluation of the most informative regions Figure 7.3 shows the most relevant region in darker shades. It seems that inferior part of the tibia bone is the region more relevant for prognosis of cartilage loss in this population. For 3, 6 and 12 regions (**a**, **b** and **d** in the figure) the inferior part has clearly the high percentages of relevance. For 8 regions (**c** in the figure), this relevance is less clear since the inferior part of the tibia mixed with the region under the subchondral bone. For 24 regions (**e** in the figure) the inferior and the medial region close to the cortical bone were selected as more relevant. For 45 regions the medial region close to the cortical bone is also selected as more relevant. However, considering the prognosis evaluation for 24 and 45 regions, the relevance of the medial region close to the cortical bone may be inconclusive.

Once identified the most informative texture region, for the next results (subsections 7.4, 7.4, 7.4), the bone marker was defined based on the outcome of mi-SVM considering that each bag had only two instances: one instance corresponded to the inferior part of the tibia bone and the other instance corresponded to the rest of the ROI.
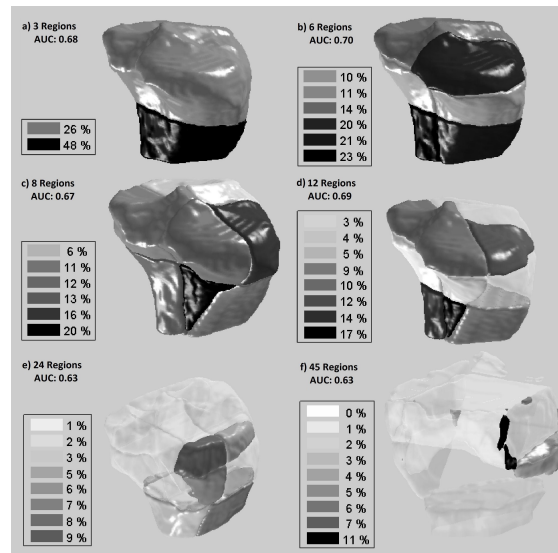
Figure 7.3: The figure shows a ROI divided in 3, 6, 8, 12, 24 and 45 regions, respectively in a, b, c, d, e and f. The images corresponds to left knees and the percentages correspond to how many times each region was selected as witness in the mi-SVM.

C. Confounding factors analysis   The results of the confounding factors analysis are shown in Table 7.2, where the OR's represent the ability of age, gender and BMI of predicting cartilage loss (second and third columns), along with their relation to the bone marker (fourth and fifth columns).

The bone marker based on the MIL framework predicted the rapid/slow progressors with OR of 3.7, with CI 2.2 – 6.0 and AUC of 0.72 (p-value < 0.0001). After linear correction for age, gender and BMI, the OR was 3.2 (CI 1.9 – 5.3) and the AUC was 0.68 (p-value 0.0001).

D. Tertiles segregation   We also evaluated the classification of progressors by tertiles. The bone structure marker could segregate patients experiencing the greatest risk of cartilage loss. The OR for the prediction of the first tertile against the third one was OR 8.3 (CI 4.3 – 15.9) with AUC of 0.82 prior to correction and OR 5.3 (CI 2.8 – 9.9), AUC of 0.77 after normalization in relation to the confounding factors. Figure 7.4 illustrates the cartilage loss

Table 7.2: Confounding factors analysis for prognosis of cartilage loss.

| Factor | Cartilage loss | | Bone marker | |
	OR	CI	OR	CI
Age	0.64	0.4−1.0	0.76	0.5−1.2
Gender	1.72	1.1−2.8	2.07	1.3−3.4
BMI	1.67	1.0−2.7	1.88	1.6−3.1

separated out by above/below the median of the bone marker and by tertiles.

E. Relation between the texture analysis outcome and pathological features   Table 7.3 shows the OR for the relations between the radiograph features evaluated and the cartilage loss, along with the relations between the features and the bone structure marker. Although not significant, the vertical trabecularization presented the highest relation to cartilage loss and to the bone marker. As a remark, the segregation between slow/rapid progressors in the 30 same samples using the bone structure marker had OR of 42.2 (CI 5.1 − 346.9).

Table 7.3: Relations between pathological features and cartilage loss and the bone structure marker

| Measure | Cartilage loss | | Bone marker | |
	OR	CI	OR	CI
MJSN	1.00	0.2− 4.6	1.00	0.2− 4.6
OP	1.83	0.7− 9.3	1.83	0.4− 9.3
VTRAB	7.00	1.1−46.0	2.36	0.5−12.4
MPHYT	1.75	0.4− 8.0	1.00	0.2− 4.5

MJSN: medial joint space narrowing, OP: osteopenia,
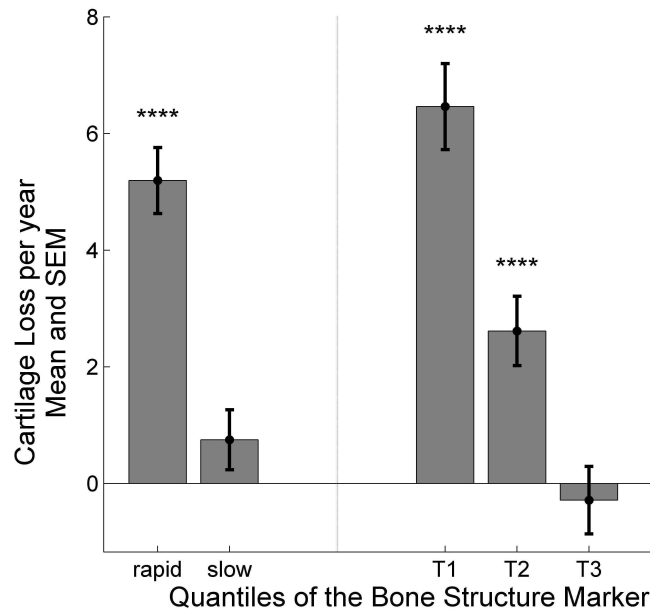VTRAB: vertical trabecularization, MPHYT: medial osteophyte

Figure 7.4: Cartilage volume losses separated out by above/below the median of the bone marker (5.2% and 0.75% of cartilage loss) on the left of the dotted line. The tertiles are on the right side of the dotted line with losses of 6.5, 2.6 and -0.3% for first, second and third tertiles respectively. The mean of the measurements are shown (with bars illustrating the standard error of the mean, SEM). The level of statistically significant separation was marked by stars: four stars correspond to p < 0.0001.

7.5   Discussion

For prediction of rapid/slow progressors of cartilage loss, the outcome showed that the presented MIL texture analysis framework reached a prognosis ability comparable to previous results [102], with the advantage of allowing the investigation of which region of the ROI are more related to cartilage loss. The MIL approach had a performance of OR 3.7, while previously we had OR of 3.9, on the same population.

The confounding factors values were aligned with recent prospective studies demonstrating obesity as a primary risk factor for incident knee OA [93], along with studies showing that women have increased rates of cartilage loss and progression of cartilage defects at the knee than men [127].

Moreover, the mean of cartilage loss for first, second and third tertiles was 6.5, 2.6 and -0.3% respectively. Patients from the first tertile are of particular interest as they are most relevant to include in clinical trials. Additionally, they are likely to have the worst prognosis with greater risk of surgical intervention by joint replacement [110].

Comparing mi-SVM and MILES, the mi-SVM approach presented a better performance for 5 out of 7 evaluations. For 3, 6, 8 and 12 regions it reached AUC around 0.70. Contrary to the expectation, MILES approach had a comparable performance evaluation only when analysing the whole ROI as the unique instance of a bag, which cannot support the investigation of the most relevant region.

When analysing which region of the tibial trabecular bone was more related to cartilage loss, the results showed the inferior part of the tibial bone was classified as the most relevant. The region beneath the subchondral area was surprisingly evaluated as less important. Furthermore, the medial part of the tibia bone was clearly more relevant than the lateral area. This outcome was not remarkable, since the experiment evaluated the relation to the medial compartment of the cartilage.

There is some indication that anatomical features related to cartilage loss affect differently subregions of the bone [24,126]. Despite the non-conclusive results, the investigation of the most relevant region brought us the question of which pathological features the bone marker could be capturing to define the risk of cartilage loss in the selected region. The outcome of our experiments (see Table 7.3) reinforced a previous hypothesis [102] of a slight relation of cartilage loss to vertical trabecularization. Unfavourably, it was not possible to achieve a high confidence level in these findings due to the small number of samples. Therefore, the results suggests further investigation on this hypothesis.

We performed a preliminary analysis of the correlation between the bone marker and some biochemical markers including tibial MRI cartilage markers of thickness, smoothness, curvature, homogeneity, among others. For simplicity this evaluation was not included in the Results section; nevertheless, it is interesting to note the correlation between the bone marker and medial tibial cartilage homogeneity had a significant result (p-value < 0.001) of OR 2.49 (CI 1.5 – 4.1). Recently, it was shown that cartilage homogeneity

quantified by MRI may capture the biochemical changes undergoing in the cartilage and could be a potential marker for early detection of knee OA [115]. We suggest a further investigation combining cartilage homogeneity and our bone marker, the combined marker could potentially increase the prediction ability of cartilage loss or early OA.

In this study, the relatively small number of subjects suggest that the findings need to be validated on other populations. Specifically, the statistical analysis on the relevant region would be more conclusive if validated in another data set. Furthermore, we suggest validation of our findings on high-field MRI scans, which may allow a more accurate and precise cartilage volume measurement and tibial trabecular bone quantification.

7.6 Conclusion

The applied MIL texture analysis demonstrated the feasibility to capture tissue changes and demonstrated the potential for automatically predicting the cartilage volume loss by analysing regions of tibial trabecular bone structure separately. The identification of the inferior part of the tibial bone as the most relevant region has enabled a preliminary finding that needs further investigation. The presented statistical analysis of the bone marker reinforced a previous hypothesis of a slight relation of cartilage loss to vertical trabecularization. Our results underlined the importance of the bone for prognosis of cartilage loss in OA. In practical terms, this investigation can contribute to the development of a new marker able to predict disease progression and identify patients most likely to progress.

Chapter 8

Sparse linear models on texture analysis

This chapter is based on the following manuscript. The introduction was rewritten to avoid redundancy when comparing to Chapter 1. The introductory description about PLS was omitted since it is detailed in Chapter 5. The same with the image data acquisition and feature computation, since they were detailed in Chapter 3. Apart from these changes, the contents are similar.

J. Marques, L. Clemmensen, E. B. Dam, "Diagnosis and prognosis of osteoarthritis by texture analysis using sparse linear models", Proceedings of the Workshop on Sparsity Techniques in Medical Imaging (STMI) in conjunction with 15th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI).

The increment of dimensionality in data sets leads to a loss of meaning for each feature and possibly decreases the accuracy of the model [21]. In such situations, sparse methods can reduce the non relevant features by adding an appropriate penalty term to the objective function. The induced sparsity has the potential of yielding a simplified and more interpretable model of the scientific problem been investigated.

Different forms of the penalty terms have been proposed in the literature

[54]. The Ridge regression [66] minimizes a penalized objective function by adding an L2 penalty term and the LASSO [135] adds an L1 penalty to the objective function. The Elastic Net combines L1 and L2 penalty, selecting some variables using LASSO and shrinking some variables according to Ridge. The resulting method can be seen as a variable selection strategy, since some of the estimated variables are forced to zero, depending on the size of this penalty.

In this chapter, we used the texture analysis framework presented in Section 5.2 to compare the performance of the DR strategies using a PLS and LDA methods with and without a sparse approach.

8.1   Background

Dimensionality reduction using LDA

LDA can be derived using different approaches. The Fisher LDA estimates a low-dimensional discriminative space defined by a linear transformation that maximizes the ratio of between-class scatter to within-class scatter.

In an alternative approach, the optimal scoring implementation recasts the classification problem as a regression problem. The categorical variables are turned into quantitative variables by defining $\mathbf{Y}$ as an $n{\times}j$ matrix of dummy variables for the j classes and n observations. By linear regression, the algorithm assigns scores to the classes, where the coefficient matrix reflect the optimal scores. See more details in [65].

Though LDA often performs quite well in low-dimensional data, it is known to fail when the number of features is larger than the number of observations [26]. In this case, LDA cannot be applied directly, without some regularization.

To deal with the high dimensions, the sparse LDA presented in [26] applies an Elastic Net penalty to the coefficient vectors in the optimal scoring interpretation of LDA. Besides performing classification and feature selection simultaneously, the imposed sparseness criterion of this approach allows to set the exact number of non-zero loadings desired in each discriminative direction.

Dimensionality reduction using PLS

The PLS regression suggests the use of supervised dimension reduction by considering both the feature matrix **X** and the response variable **Y**. This regression returns linear combinations of the features, called latent factors or **X**-scores, which are used to predict the transformed response variable (**Y**-scores). The **X**-scores are chosen so that the relationship between successive pairs of scores is as strong as possible (See Section 5.1.2 for more details).

A large number of noise features can force the PLS loadings to divert from the direction that relates **X** and **Y**, which can cause inconsistency [73]. Considering that it can attenuate estimates of the regression parameters, Hyonho et al. [73] proposed a PLS formulation with imposed sparsity on the direction vectors. The proposed sparse method is equivalent to the Elastic Net approach, which selects some variables and shrinks some values towards zero.

Since the PLS sparse approach tends to avoid inconsistency on the direction vectors, in the present work, we investigated whether it can efficiently identify the relevant texture features for diagnosis of OA and prognosis of cartilage loss. Building on Hyonho et al.'s proposed formulation we implemented a sparse PLS algorithm and included it as a feature selection step to the texture analysis framework.

8.2 Application of the framework for OA diagnosis and prognosis

In this section, we briefly describe the texture analysis framework and its application to diagnosis and prognosis of OA. The overall framework included the following ordered steps: segmentation of the ROI, features computation, dimensionality reduction, classification and evaluation. Section 3.1 introduces a brief description of the data collection, Section 3.3 explains how the feature were computed and Section 8.2.2 explains the dimensionality reduction method including the detailed implementation of sparse PLS.
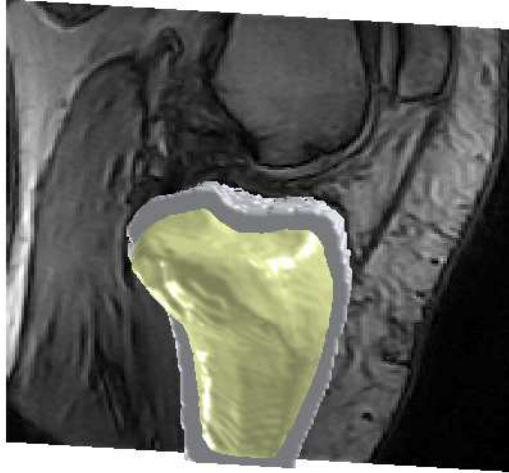
Figure 8.1: a) Automatically segmented tibia bone in light gray and the region-of-interest (the trabecular bone) in gold.

8.2.1 ROI definition

The ROI definition followed the same approach described in Section 5.3.1. The only difference refers to the morphological erosion. The experiments described in this chapter applied a erosion of approximately 5 mm to remove the cortical bone. Figure 8.1 shows an example of an ROI.

8.2.2 The dimensionality reduction method

As a pre-processing step, the feature set was normalized to zero mean and a standard deviation of one. Next, the sparse PLS algorithm defined the selected features and the number of PLS latent factors used in the final model.

The sparse PLS algorithm: The original implementation of the SIMPLS algorithm [39] uses conjugate gradient (CG) [128] to compute the coefficients. At each iteration, one column of the matrix **W** is computed based on the correlation of each **X**-column with the **Y** variable. The sparse PLS algorithm applied hard threshold by imposing zero values to all elements of the **W**-column with absolute value less than the specified threshold. The candidate

thresholds were defined to be logarithmically spaced between the minimum and the maximum value of the first latent factor. Using cross-validation, each candidate threshold was sent to the sparse PLS algorithm and evaluated. This intermediate evaluation considered different numbers of PLS latent factors. For the best evaluation, the algorithm identified the selected features by the non-zero value in the **W**-matrix. Note that the threshold and number of PLS latent factors were optimized by cross-validation, while the selected features were determined by the sparsity algorithm. The final model included the selected feature set and number of PLS latent factors.

The sparse LDA evaluation: To compare both approaches of sparsity, we replaced the sparse PLS by the sparse LDA. During the training phase, a cross-validation strategy optimized the number of selected features and the weight of the L2-norm for elastic net regression. The best combination of theses two parameters was used by the the sparse LDA algorithm in the simultaneous feature selection and classification step of the final model.

8.2.3 Classification and Evaluation

The performance of the methods was evaluated using a 10-fold, cross-validation approach. For classification, the framework used the Fisher LDA. For evaluation, we measured the AUC.

8.3 Experiments and Results

To investigate the performance of the sparsity methods on the model accuracy, we performed five experiments in each dataset. The experiments applied different DR methods to the 534 original features generated in accordance with Section 3.3.

The first experiment applied the Fisher LDA considering the Moore-Penrose pseudo-inverse [9] of the covariance matrix, since the number of samples was less than the number of the original features. The second experiment applied the SIMPLS algorithm of PLS regression. The DR step of the training phase defined the number of PLS latent factors to use in the final model. The next experiment evaluated the performance of the PLS forward

Table 8.1: Diagnosis evaluations of the different methods: Fisher LDA, PLS regression, PLS with forward feature selection (PLS-FFS), sparse PLS and sparse LDA. The third column shows the number of features used in each CV fold evaluation. The fourth column shows the number of features used across all CV folds evaluations.

	Diagnosis		
Method	AUC	features per fold	features
LDA	0.86	534	534
PLS	0.88	534	534
PLS-FFS	0.89	116	212
SPLS	0.93	349	500
SLDA	0.89	113	268

Table 8.2: Prognosis evaluations of the different sparse and non-sparse methods.

	Prognosis		
Method	AUC	features per fold	features all folds
LDA	0.63	534	534
PLS	0.67	534	534
PLS-FFS	0.69	43	108
SPLS	0.70	44	56
SLDA	0.59	136	306

feature selection method presented in Section 5.2.1. Likewise, the two last experiments evaluated the performance of the sparse PLS and the sparse LDA. Their outcomes were propagated to the final classification process. Tables 8.1 and 8.2 compare all the evaluations.

8.4 Discussion and Conclusion

Comparing LDA and sparse LDA, the results indicated that sparse LDA improved the diagnosis evaluation and reduced considerably the final feature

space. But contrary to expectations, there was no detectable improvement to the prognosis evaluations. One possible explanation can be overfitting, since during the training phase, the method had median AUC of 0.99 for both datasets.

However, by including sparsity in the PLS algorithm we could increase the model accuracy and identify the subset of features actually used by the texture analysis framework. In general, the sparse PLS performed better than all other evaluated methods. The accuracy improvement was more expressive in the diagnosis evaluation, where the AUC reached 0.93. Comparatively, a recent study analysing a linear combination of morphometric and structural cartilage markers in the same population scored AUC of 0.84 [34]. Although the studies analysed different anatomical structures, the results showed the sparse PLS captured the texture changes and had diagnostic ability superior to other biomarkers of OA.

In the prognosis, the sparse PLS reached an AUC of 0.70. Although the performance was only slightly better than the other methods, the number of features selected were only 9% of the available ones. Considering all cross validation folds, we can notice some overlap between the selected feature sets, indicating the stability of the model. The sparse PLS decreased the model complexity, which can potentially contribute to a better understanding of the anatomical characteristics of the data being analysed. However, the comparison of these methods using another dataset is key for verifying the results.

In conclusion, we presented a investigation of sparsity methods for dimensionality reduction in texture analysis. The results illustrated that by including a sparsity approach, our framework limited the number of features used by the final model and increased the performance ability of separating healthy and OA subjects.

Chapter 9

Summary and general discussion

9.1 Summary

The work presented in this thesis investigated the feasibility of quantifying OA by analysis of the trabecular bone structure in knee MRI. Chapter 4 presented the first experiments for diagnosis of OA, where a previous version of our framework [103] evaluated the subchondral medial trabecular tibial bone. The ROI was contained within the medial tibial condyle, below the central, load-bearing cartilage, in an area expected to have relatively homogeneous biomechanical stress. The results showed that the presence of OA could be quantified by a bone structure marker. We also presented an investigation of aggregate markers based on measurements targeting different anatomical structures including medial tibial cartilage and subchondral tibial bone.

A new texture analysis framework was presented in Chapter 5. Different linear feature selection approaches were evaluated and the new methodology included a robust PLS-based dimensionality reduction method, since it demonstrated better consistency and stability in selecting the features.

In Chapter 6, the new framework was evaluated in a longitudinal study where MRI scans of knees was used to quantify the tibial trabecular bone in a bone marker for OA diagnosis and another marker for prediction of tibial cartilage loss. For this population, the bone markers demonstrated to be superior to other biomarkers of OA. Besides it, a preliminary radiological reading of the high and low risk knees slightly suggested that the prognosis

bone marker could be capturing aspects of the vertical trabecularization of the tibial bone to define the risk of cartilage loss.

Chapter 7 described a methodology to identify the most informative region of a ROI and showed which region of the tibial bone was more related to cartilage loss. The structure of the trabecular tibial bone was divided in localized subregions in an attempt to capture the different pathological features occurring at each location. We applied multiple-instance learning, where each subregion was defined to be one instance and a bag held all instances over a full region-of-interest. Aiming to potentially assimilate the structure of the trabecular bone anatomy in each region, we quantified subregions of bone separately. As a rule, if the learning algorithm identified some tissue change related to cartilage loss in at least one region (instance) of the whole ROI, the subject was classified as rapid progressor of cartilage loss. The results showed that the inferior part of the tibial bone was classified as the most relevant region for prognosis of cartilage loss and the preliminary radiological reading of a subset of the samples suggested the bone marker also captured aspects of the vertical trabecularization to define the most relevant region.

An investigation into sparsity methods for dimensionality reduction in texture analysis was presented in Chapter 8. By including a sparsity approach, our framework limited the number of features used by the final model and increased the performance ability of separating healthy and OA subjects.

Finally, this last chapter concludes the dissertation by summarizing the thesis and providing a general discussion of its content.

9.2 Discussion

Framework

This dissertation presented a texture analysis framework that quantifies a specific region of interest and produces a marker according to a pre trained relation. All the process, from segmentation to marker assessment, is fully automatic, which allows scalability to large, multi-center studies and could be used, for example, in pharmaceutical studies or in clinical research with

large sets of data.

Another advantage of the presented methodology was the use of low cost MRI equipments. The results revealed that tissue changes and disease-related bone structure characteristics were reflected in the low-field MRI appearance, even though smaller individual structures, like the trabecular bone, were not visible.

The proposed methodology might be applicable to quantify other ROI's related to different abnormalities. A contributing factor for this potential adaptability is the generic set of texture descriptors extracted from the images. A broad representation of the image can lead to a better discriminative ability. The applied Gaussian derivative filters allowed to capture the local structure of the voxels and by combining it with a multi-scale strategy, the extracted information was extended to scales larger than the voxel resolution. On the basis of this, the proposed framework can potentially be applied to analyse pathological features of different sizes.

OA quantification

In this project, the framework was validated for diagnosis and prognosis of cartilage loss by quantifying the tibial trabecular bone. The initial experiments (Chapter 4) evaluated the tibial subchondral bone area and the subsequent experiments evaluated the whole tibial trabecular bone (see Figure 9.1). The ROI expansion was implemented after the generalization of a voxel classification algorithm that allowed the segmentation of the tibial bone. A comparison between Table 4.1 and Table 8.1 can indicate the influence of the ROI expansion in the results of the leave-one-out LDA, the tibial subchondral bone analysis reached generalization AUC of 0.82 while the whole tibial bone had AUC of 0.86.

In addition to this enhancement, the proposed improvements to the framework increased the diagnostic ability to AUC 0.92 when using the PLS-FFS method (described in Chapter 5) and to 0.93 when using the SPLS (described in Chapter 8). These results outperformed structural cartilage markers and biochemical markers present in the literature (see Chapter 4).

In practical terms, the developed quantification can contribute to the development of a new OA marker able to predict disease progression and

Figure 9.1: a) The figure shows in dark gray the tibial subchondral trabecular bone, which corresponds to the ROI used in Chapter 4; and in light gray the tibial trabecular bone, used in the experiments of the remaining chapters.

identify patients most likely to progress. From a theoretical point of view, our findings underlined the importance of the bone to the OA diagnosis, prognosis of cartilage loss and understanding the disease processes. It is aligned with the gradual shift on the characterization of OA from a cartilage centred view towards the whole joint system approach.

Dimensionality reduction

To achieve a satisfactory discrimination ability, the framework should be able to identify which of the extracted features are best representing the relation between the ROI structure to the problem being investigated. A less complex

model can potentially contribute to a better understanding of the anatomical characteristics of the data.

We presented different dimensionality reduction methods for identifying the most relevant features. Chapter 4 compared different linear feature selection methods and Chapter 8 compared some alternative approaches implementing sparsity. The method with best classification ability was the sparse PLS, both for prognosis and diagnosis. For the prognosis evaluation, the sparse PLS reduced considerably the number of features and indicated good stability. However, in the diagnosis, the method was more unstable and less effective. This might be due to overfitting during the training phase. Considering feature selection efficacy and accuracy performance together, the method that applies PLS followed by forward feature selection performed satisfactorily in both problems.

Machine learning approaches

Another contribution of this investigation was the application of multiple instances learning techniques to identify which bone regions was potentially more suitable for prognosis of cartilage loss. The classification identified the inferior part of the tibial bone as the most relevant region (see Chapter 7). This result was surprising, intuitively we would expect some region closer to the tibial cartilage. The relation of the bone marker to the vertical trabecularization might bring some clarification to this outcome; however, this finding merits further investigation.

Is interesting to note that two conceptually different learning approaches were evaluated. The first approach, applied in the investigations described from Chapter 4 to Chapter 6, classifies the images by analysing the whole ROIs at once, while the second approach, the MIL described in Chapter 7, classifies the images by analysing regions of the ROIs separately. In the MIL approach, the classification of the image can be based on the analysis of only a small part of the ROI. Intuitively, if the abnormality being investigated can affect only part of the ROI, this approach can bring improvement to the classification performance. The results in Chapter 7 do not present better performance than the previous ones. Based on this, one could argue that the whole tibial trabecular bone presents some abnormality leading to cartilage

loss, having some regions more affected than others. The outcome of this analysis warrants further investigation, by applying the method in a different population and also by including more radiographic information about the images.

Relation to pathological features

Chapters 6 and 7 presented a preliminary finding suggesting a slight relation of the developed prognosis bone quantification marker to vertical trabecularization. This analysis relied on 25 samples and had information of four pathological features, vertical trabecularization, medial osteophyte and osteopenia medial along with joint space narrowing. Due to the small number of samples, it was not possible to achieve a high confidence level for this analysis, suggesting further investigation on this relation. Another clinical challenge that remains for future research is an examination on the relationship between the selected feature sets and the pathological features they can represent, e.g., BMLs, osteophytes, bone erosions, subchondral cysts, bone attrition, among others. For this analysis, we would need not only the labels but also the information of which pathological features are present in the images and where they are localized.

Aggregate markers

Chapter 4 presented an investigation of aggregate markers based on measurements targeting different anatomical structures. The results demonstrated that measurements from MRI and measurements from biochemical markers are complementary and aggregate markers can be superior than the individuals ones.

9.3 Limitations and future work

The high percentage of the population (80%) in early stages of OA suggest that the conclusions are more accurate for progression during the early stages of OA. It would be interesting to evaluate our findings for later stages of OA as well, considering other population with different stages distribution.

Due to the relatively small number of subjects, the statistical analysis on both the most relevant region and on the pathological features been captured by our bone marker would be more conclusive if the results can be reproducible in other, and maybe larger, data sets.

Much more research is needed to determine the clinical features that the bone structure marker describe. Besides the pathological features evaluated, potential candidates could include BMLs, osteophytes, bone erosions, subchondral cysts, bone attrition, among others. It would be also interesting to examine the the relationship between one or a group of selected features and the pathological features they can represent.

Furthermore, more research is required to integrate cartilage and bone into a comprehensive approach that can increase our understanding of the disease processes. For example, we suggest an investigation combining cartilage homogeneity and our prognosis bone marker, the combined marker could potentially increase the prediction ability of cartilage loss. Additionally, could be interesting re-evaluate the aggregate marker presented in Chapter 4 considering the bone diagnostic marker based on the improved framework (Chapter 5), analysing the whole tibial bone.

Additionally, we suggest validation of our findings on high-field MRI scans, which may allow a more accurate and precise cartilage volume measurement and tibial trabecular bone quantification.

Use of the framework for clinical trials or in clinical practice

Before considering the use of the framework for clinical trials or in clinical practice, we suggest some extra validation. For example, we suggest validation of our findings on a different study population. This could express the variability of the framework quantification under different equipment, OA stages distribution, calibration criteria, imaging protocol, among others.

The MRI images evaluated in this project was obtained in an one-center study. It would interesting to evaluate the reproducibility of the framework in a multi-center study, where most of the methodology remains the same, but some variety is included in the image acquisition procedure.

The requirements for using the framework for clinical practice is more elaborate than for clinical trials. The supervised learning approaches im-

plemented in the framework requires the specification of certain parameters prior the quantification. Besides it, slight differences in the device settings might introduce not expected measurements and generate an image improper for quantification.

Despite the fact that ideally, the quantification should be made using settings and equipment which have been tested in advance, ensuring reliability, the challenge is to specify parameters that reflects all the complexity and variability in a clinic image acquisition procedure. To deal with this limitation, further investigation on transfer learning techniques [108] might bring some clarification.

As initial steps for the implementation in clinical practice, the framework should be trained in a large population that includes all the OA stages and at least some of the variability in the image acquisition procedure. Afterwards, all the limitations and requirements should be specified in a documentation that includes precise procedures for subject preparation, image acquisition and general guidelines. Additionally, the evaluation of the framework reproducibility becomes crucial, since the image quantification made in one clinic should be consistent with those elsewhere.

9.4   Conclusion

This thesis presented a quantification method with demonstrated potential for capturing bone tissue changes and a strategy to evaluate the more affected region of the ROI. In clinical terms, besides presenting an OA marker with diagnostic ability superior to other biomarkers of OA, our findings contributed to the development of a new bone marker able to identify patients most likely to progress, underlining the importance of the trabecular bone to the understanding of the disease process.

Bibliography

[1] Herve Abdi. Partial least squares (pls) regression. *Encyclopedia of social sciences research methods ed*, 21(4):1–7, 2003.

[2] Ethem Alpaydin. *Introduction to Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2004.

[3] B K Alsberg, D B Kell, and R Goodacre. Variable selection in discriminant partial least-squares analysis. *Analytical Chemistry*, 70(19):4126–4133, 1998.

[4] R D Altman, J F Fries, D A Bloch, J Carstens, T D Cooke, H Genant, P Gofton, H Groth, D J McShane, and W A Murphy. Radiographic assessment of progression in osteoarthritis. *Arthritis Rheum*, 30(11):1214–25, 1987.

[5] Stuart Andrews, Ioannis Tsochantaridis, and Thomas Hofmann. Support vector machines for multiple-instance learning. In *Advances in Neural Information Processing Systems 15*, pages 561–568. MIT Press, 2003.

[6] Suchandrima Banerjee, Roland Krug, Sven Prevrhal, and Sharmila Majumdar. New imaging techniques for bone. In Robert A. Adler, editor, *Osteoporosis*, Contemporary Endocrinology, pages 51–76. Humana Press, 2010.

[7] Matthew Barker and William Rayens. Partial least squares for discrimination. *Journal of Chemometrics*, 17(3):166–173, 2003.

[8] Anne-Christine Bay-Jensen, Suzi Hoegh-Madsen, Erik Dam, Kim Henriksen, BodilCecillie Sondergaard, Philippe Pastoureau, Per Qvist, and MortenA. Karsdal. Which elements are involved in reversible and irreversible cartilage degradation in osteoarthritis? *Rheumatology International*, 30:435–442, 2010.

[9] A. Ben-Israel and T.N.E. Greville. *Generalized Inverses: Theory and Applications*. CMS Books in Mathematics. Springer, 2003.

[10] Kevin S. Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft. When is "nearest neighbor" meaningful? In *Proceedings of the 7th International Conference on Database Theory*, ICDT '99, pages 217–235, London, UK, 1999. Springer-Verlag.

[11] Jinbo Bi, Kristin P. Bennett, Mark Embrechts, Curt M. Breneman, Minghu Song, Isabelle Guyon, and André Elisseeff. Dimensionality reduction via sparse support vector machines. *Journal of Machine Learning Research*, 3:2003, 2003.

[12] Jinbo Bi and Jianming Liang. Multiple instance learning of pulmonary embolism detection with geodesic distance along vascular structure. In *CVPR*. IEEE Computer Society, 2007.

[13] G. Blumenkrantz, C. T. Lindsey, T. C. Dunn, H. Jin, M. D. Ries, T. M. Link, L. S. Steinbach, and S. Majumdar. A pilot, two-year longitudinal study of the interrelationship between trabecular bone and articular cartilage in the osteoarthritic knee. *Osteoarthritis and Cartilage*, 12(12):997–1005, December 2004.

[14] C Boileau, J Martel-Pelletier, F Abram, J-P Raynauld, E Troncy, M-A D'Anjou, M Moreau, and J-P Pelletier. Magnetic resonance imaging can accurately assess the long-term progression of knee structural changes in experimental dog osteoarthritis. *Ann Rheum Dis*, 67(7):926–32, 2008.

[15] K. Vanden Branden and M. Hubert. Robustness properties of a robust partial least squares regression method. *Analytica Chimica Acta*, 515(1):229 – 241, 2004.

[16] K.D. Brandt, P. Dieppe, and E.L. Radin. Etiopathogenesis of osteoarthritis. *Rheum Dis Clin North Am*, 34(3):531–59, 2008.

[17] Kenneth D. Brandt, Michael Doherty, and L. Stefan Lohmander. *Osteoarthritis*. New York: Oxford University, second edition, 2006.

[18] Kenneth Bryan, Lorraine Brennan, and Padraig Cunningham. Metafind: A feature analysis tool for metabolomics data. *BMC Bioinformatics*, 9(1):470, 2008.

[19] J. C. Buckland-Wright, J. A. Lynch, and D. G. Macfarlane. Fractal signature analysis measures cancellous bone organisation in macroradiographs of patients with knee osteoarthritis. *Annals of the Rheumatic Diseases*, 55(10):749–755, 1996.

[20] Prabir Burman. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*, 76(3):503–514, 1989.

[21] P S P WangEditor C H Chen L F Pau, editor. *Texture Analysis, In The Handbook of Pattern Recognition and Computer Vision (2nd Edition)*, pages 207–248. World Scientific Publishing Co., 1998.

[22] Yixin Chen, Jinbo Bi, and James Z. Wang. MILES: Multiple-Instance Learning via Embedded Instance Selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):1931–1947, 2006.

[23] K. Chernoff and M. Nielsen. Weighting of the k-Nearest-Neighbors. *International Conference of Pattern Recognition (ICPR)*, 2010.

[24] K. Chiba, M. Uetani, Y. Kido, M. Ito, N. Okazaki, K. Taguchi, and H. Shindo. Osteoporotic changes of subchondral trabecular bone in osteoarthritis of the knee: a 3-t mri study. *Osteoporosis International*, 23:589–597, 2012.

[25] Il-Gyo Chong and Chi-Hyuck Jun. Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory System*, 78(1-2):103–112, June 2005.

[26] Line Clemmensen, Trevor Hastie, Daniela Witten, and Bjarne Ersbøll. Sparse discriminant analysis. *Technometrics*, 53(4):406–413, 2011.

[27] D Coggon, I Reading, P Croft, M McLaren, D Barrett, and C Cooper. Knee osteoarthritis and obesity. *International Journal of Obesity - Related Metabolic Disorders*, 25(5):622, 2001.

[28] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[29] Massoud Daheshia and Jian Yao. The bone marrow lesion in osteoarthritis. *Rheumatology International*, 31:143–148, 2011. 10.1007/s00296-010-1454-x.

[30] Enrico Dall'Ara, Caroline Öhman, Massimiliano Baleani, and Marco Viceconti. Reduced tissue hardness of trabecular bone is associated with severe osteoarthritis. *Journal of Biomechanics*, 44:1593–1598, 2011.

[31] E B Dam, I Byrjalsen, M A Karsdal, P Qvist, and C Christiansen. Increased urinary excretion of c-telopeptides of type ii collagen (ctx-ii) predicts cartilage loss over 21 months by mri. *Osteoarthritis and Cartilage*, 17(3):384–389, 2009.

[32] E. B. Dam, J. Folkesson, P. C. Pettersen, and C. Christiansen. Automatic cartilage thickness quantification using a statistical shape model. In *MICCAI Joint Disease Workshop*, pages 42–49, 2006.

[33] E. B. Dam, J. Folkesson, P. C. Pettersen, and C. Christiansen. Automatic morphometric cartilage quantification in the medial tibial plateau from MRI for osteoarthritis grading. *Osteoarthritis and Cartilage*, 15(7):808 – 818, 2007.

[34] E. B. Dam, M. Loog, C. Christiansen, I. Byrjalsen, J. Folkesson, M. Nielsen, A. Qazi, P. C. Pettersen, P. Garnero, and M. A. Karsdal. Identification of progressors in osteoarthritis by combining biochemical and MRI-based markers. *Arthritis Research & Therapy*, 11(4):R115, 2009.

[35] Erik B. Dam, P. Thomas Fletcher, and Stephen M. Pizer. Automatic shape model building based on principal geodesic analysis bootstrapping. *Medical Image Analysis*, 12(2):136–151, 2008.

[36] Erik B Dam, Marco Loog, Claus Christiansen, Inger Byrjalsen, Jenny Folkesson, Mads Nielsen, Arish A Qazi, Paola C Pettersen, Patrick Garnero, and Morten A Karsdal. Identification of progressors in osteoarthritis by combining biochemical and mri-based markers. *Arthritis Res Ther*, 11(4):R115, 2009.

[37] M. Daszykowski, Y. Vander Heyden, and B. Walczak. Robust partial least squares model for prediction of green tea antioxidant capacity from chromatograms. *Journal of Chromatography A*, 1176(1-2):12 – 18, 2007.

[38] M. Daszykowski, K. Kaczmarek, Y. Vander Heyden, and B. Walczak. Robust statistics in data analysis - a review: Basic concepts. *Chemometrics and Intelligent Laboratory Systems*, 85(2):203 – 219, 2007.

[39] S. de Jong. Simpls: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3):251–263, March 1993.

[40] D K Dedrick, S A Goldstein, K D Brandt, B L O'Connor, R W Goulet, and M Albrecht. A longitudinal study of subchondral plate and trabecular bone in cruciate-deficient dogs with osteoarthritis followed up for 54 months. *Arthritis Rheum*, 36(10):1460–7, 1993.

[41] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3):837–845, 1988.

[42] Thomas G. Dietterich, Richard H. Lathrop, and Tomás Lozano-Pérez. Solving the multiple instance problem with axis-parallel rectangles. *Artif. Intell.*, 89(1-2):31–71, 1997.

[43] Jianrui Ding, H.D. Cheng, Jianhua Huang, Jiafeng Liu, and Yingtao Zhang. Breast ultrasound image classification based on multiple-instance learning. *Journal of Digital Imaging*, 25:620–627, 2012.

[44] Murat Dundar, Glenn Fung, Balaji Krishnapuram, and R. Bharat Rao. Multiple-instance learning algorithms for computer-aided detection. *IEEE Trans. Biomed. Engineering*, 55(3):1015–1021, 2008.

[45] F. Eckstein, D. Burstein, and T. M. Link. Quantitative MRI of cartilage and bone: degenerative changes in osteoarthritis. *NMR in Biomedicine*, 19:822–854, 2006.

[46] Felix Eckstein, Wolfgang Wirth, Martin I Hudelmaier, Susanne Maschek, Wolfgang Hitzl, Bradley T Wyman, Michael Nevitt, Marie-Pierre Hellio Le Graverand, and David Hunter. Relationship of compartment-specific structural knee status at baseline with change in cartilage morphology: a prospective observational study using data from the osteoarthritis initiative. *Arthritis Res Ther*, 11(3):R90, 2009.

[47] B J Ejbjerg, E Narvestad, S Jacobsen, H S Thomsen, and M Østergaard. Optimised, low cost, low field dedicated extremity mri is highly specific and sensitive for synovitis and bone erosions in rheumatoid arthritis wrist and finger joints: comparison with conventional high field mri and radiography. *Ann Rheum Dis*, 64(9):1280–7, 2005.

[48] Martin Englund, Ali Guermazi, and Stefan L Lohmander. The role of the meniscus in knee osteoarthritis: a cause or consequence? *Radiol Clin North Am*, 47(4):703–12, 2009.

[49] L. Eriksson, E. Johansson, N. Kettaneh-Wold, and S. Wold. *Multi- and megavariate data analysis: Basic principles and applications*. Number 1 in Training in multivariate technology. Umetrics AB, 2006.

[50] Lindgren F, Geladi P, Berglund, Sjostrom M, and Wold S. Interactive variable selection (ivs) for pls. part 1: Theory and algorithms. *Journal of Chemometrics*, 8:349–363, 1994.

[51] D. T. Felson, R. C. Lawrence, P. A. Dieppe, R. Hirsch, C. G. Helmick, J. M. Jordan, R. S. Kington, N. E. Lane, M. C. Nevitt, Y. Zhang, M. Sowers, T. Mcalindon, T. D. Spector, A. R. Poole, S. Z. Yanovski, G. Ateshian, L. Sharma, J. A. Buckwalter, K. D. Brandt, and J. F. Fries. Osteoarthritis: New insights. part 1: the disease and its risk factors. *Annals of Internal Medicine*, 133(8):635 – 646, 2000.

[52] D T Felson, A Naimark, J Anderson, L Kazis, W Castelli, and R F Meenan. The prevalence of knee osteoarthritis in the elderly. the framingham osteoarthritis study. *Arthritis Rheum*, 30(8):914–8, 1987.

[53] D. T. Felson and T. Neogi. Osteoarthritis: is it a disease of cartilage or of bone? *Arthritis & Rheumatism*, 50(2):341–344, February 2004.

[54] Peter Filzmoser, Moritz Gschwandtner, and Valentin Todorov. Review of sparse methods in regression and classification with application to chemometrics. *Journal of Chemometrics*, 26(3-4):42–51, 2012.

[55] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. The Gaussian scale-space paradigm and the multiscale local jet. *International Journal of Computer Vision*, 18(1):61–75, April 1996.

[56] J. Folkesson, E. B. Dam, O. F. Olsen, M. A. Karsdal, P. C. Pettersen, and C. Christiansen. Automatic quantification of local and global articular cartilage surface curvature: Biomarkers for osteoarthritis? *Magnetic Resonance in Medicine*, 59(6):1340–1346, 2008.

[57] J. Folkesson, E. B. Dam, O. F. Olsen, P. C. Pettersen, and C. Christiansen. Segmenting articular cartilage automatically using a voxel classification approach. *IEEE Transactions on Medical Imaging*, 26:106–115, 2007.

[58] Cesare Furlanello, Maria Serafini, Stefano Merler, and Giuseppe Jurman. Entropy-based gene ranking without selection bias for the predictive classification of microarray data. *BMC Bioinformatics*, 4(1):54, 2003.

[59] Roemer FW, Guermazi A, Felson DT, Niu J, Nevitt MC, Crema MD, Lynch JA, Lewis CE, Torner J, and Zhang Y. Do baseline synovitis and effusion predict tibiofemoral cartilage loss over 30 months in subjects without radiographic osteoarthritis? results from the multicenter osteoarthritis (most) study. *Osteoarthritis and Cartilage*, 17:S210, 2009.

[60] Verheugen G. Commission directive 2005/28/ec laying down principles and guidelines for good clinical practice as regards investigational medicinal products for human use, as well as the requirements for authorization of the manufacturing or importation of such products. *Official Journal of the European Union*, 13(9):Legislation 091, 2005.

[61] M.L. Gray, F. Eckstein, C. Peterfy, L. Dahlberg, Y. Kim, and A.G. Sorensen. Toward imaging biomarkers for osteoarthritis. *Clin Orthop Relat Res*, (427 Suppl):S175–81, 2004.

[62] ; Hunter DJ; Roemer FW Guermazi A. Magnetic resonance imaging prevalence of different features of knee osteoarthritis in persons with normal knee x-rays. *Arthritis & Rheumatism*, 56:S128, 2007.

[63] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh. *Feature Extraction: Foundations and Applications*, volume 207 of *Studies in Fuzziness and Soft Computing*. Springer Berlin / Heidelberg, 2006.

[64] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, pages 389–422, 2002.

[65] Trevor Hastie, Andreas Buja, and Robert Tibshirani. Penalized discriminant analysis. *The Annals of Statistics*, 23(1):pp. 73–102, 1995.

[66] Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[67] J. Hohe, G. Ateshian, M. Reiser, K. Englmeier, and F. Eckstein. Surface size, curvature analysis, and assessment of knee joint incongruity with mri in vivo. *Magn Reson Med*, 47(3):554–61, 2002.

[68] A Hoskuldsson. Analysis of latent structures in linear models. *Journal of Chemometrics*, 17(1):630–645, 2003.

[69] D. J. Hunter, G. H. Lo, D. Gale, A. J. Grainger, A. Guermazi, and P. G. Conaghan. The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (bostonŰleeds osteoarthritis knee score). *Annals of the Rheumatic Diseases*, 67:206–211, 2008.

[70] David J Hunter. Insights from imaging on the epidemiology and pathophysiology of osteoarthritis. *Radiol Clin North Am*, 47(4):539–51, 2009.

[71] David J Hunter, Jiang Li, Michael LaValley, Doug C Bauer, Michael Nevitt, Jeroen DeGroot, Robin Poole, David Eyre, Ali Guermazi, Dan Gale, and David T Felson. Cartilage markers and their association with cartilage loss on magnetic resonance imaging in knee osteoarthritis: the boston osteoarthritis knee study. *Arthritis Res Ther*, 9(5):R108, 2007.

[72] David J Hunter, Yuqing Zhang, Jingbo Niu, Joyce Goggins, Shreyasee Amin, Michael P LaValley, Ali Guermazi, Harry Genant, Daniel Gale, and David T Felson. Increase in bone marrow lesions associated with cartilage loss: a longitudinal magnetic resonance imaging study of knee osteoarthritis. *Arthritis Rheum*, 54(5):1529–35, 2006.

[73] Chun Hyonho and Sündüz Keles. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):3 – 25, 2010.

[74] Christian Igel, Verena Heidrich-Meisner, and Tobias Glasmachers. Shark. *Journal of Machine Learning Research*, 9:993–996, 2008.

[75] Masako ITO. Recent progress in bone imaging for osteoporosis research. *Journal of bone and mineral metabolism*, 29(2):131–140, 2011.

[76] A. K. Jain and D. Zongker. Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):153–158, 1997.

[77] G Jones, Changhai Ding, F Scott, M Glisson, and F Cicuttini. Early radiographic osteoarthritis is associated with substantial changes in cartilage volume and tibial bone surface area in both males and females. *Osteoarthritis Cartilage*, 12(2):169–74, 2004.

[78] Sarang Joshi, Stephen Pizer, P. Thomas Fletcher, Paul Yushkevich, Andrew Thall, and J. S. Marron. Multiscale deformable model segmentation and

statistical shape analysis using medial descriptions. *TRANSACTIONS ON MEDICAL IMAGING*, 21(5):538–550, 2002.

[79] M. A. Karsdal, D. J. Leeming, E. B. Dam, K. Henriksen, P. Alexandersen, P. Pastoureau, R. D. Altman, and C. Christiansen. Should subchondral bone turnover be targeted when treating osteoarthritis? *Osteoarthritis and Cartilage*, 16(6):638–646, June 2008.

[80] A Kassner and R E Thornhill. Texture analysis: a review of neurologic mr imaging applications. *AJNR Am J Neuroradiol*, 31(5):809–16, 2010.

[81] J. H. Kellgren and J. S. Lawrence. Radiological assessment of osteo-arthrosis. *Annals of the Rheumatic Diseases*, 16(4):494–502, December 1957.

[82] E K Kemsley. Discriminant analysis of high-dimensional data: a comparison of principal components analysis and partial least squares data reduction methods. *Chemometrics and Intelligent Laboratory Systems*, 33(1):47–61, 1996.

[83] Namkug Kim, June-Goo Lee, Youngkyu Song, Hengjun J Kim, Jin S Yeom, and Gyunggoo Cho. Evaluation of mri resolution affecting trabecular bone parameters: Determination of acceptable resolution. *Magn Reson Med*, 2011.

[84] Jaakko Kinnunen, SÃűren Bondestam, Aarne Kivioja, Juhani Ahovuo, Sanna-Kaisa Toivakka, Ilkka Tulikoura, and Tiina Karjalainen. Diagnostic performance of low field mri in acute knee injuries. *Magnetic Resonance Imaging*, 12(8):1155 – 1160, 1994.

[85] B Kladny, K Glückert, B Swoboda, W Beyer, and G Weseloh. Comparison of low-field (0.2 tesla) and high-field (1.5 tesla) magnetic resonance imaging of the knee joint. *Arch Orthop Trauma Surg*, 114(5):281–6, 1995.

[86] P. R. Kornaat, R. Y. T. Ceulemans, H. M. Kroon, N. Riyazi, M. Kloppenburg, W. O. Carter, T. G. Woodworth, and J. L. Bloem. Knee osteoarthritis scoring system (KOSS)Ǔ inter-observer and intra-observer reproducibility of a compartment-based scoring system. *Skeletal Radiology*, 34:95–102, February 2005.

[87] Roland Krug, Andrew J. Burghardt, Sharmila Majumdar, and Thomas M. Link. High-resolution imaging techniques for the assessment of osteoporosis. *Radiologic Clinics of North America*, 48(3):601 – 621, 2010. <ce:title>Imaging of Osteoporosis</ce:title>.

[88] Uwe Kruger, Yan Zhou, Xun Wang, David Rooney, and Jillian Thompson. Robust partial least squares regression: Part i, algorithmic developments. *Journal of Chemometrics*, 22(1):1–13, 2008.

[89] W. J. Krzanowski, editor. *Principles of multivariate analysis: a user's perspective*. Oxford University Press, Inc., New York, NY, USA, 1988.

[90] Sharma L, Song J, Felson DT, Cahue S, Shamiyeh E, and Dunlop DD. The role of knee alignment in disease progression and functional decline in knee osteoarthritis. *JAMA: The Journal of the American Medical Association*, 286(2):188–195, 2001.

[91] Bier F. Lawrence JS, Bremner JM. Osteo-arthtosis. prevalence in ihe population and relationships between symptoms and x-ray changes. *Annals of the Rheumatic Diseases*, 25:l–24, 1966.

[92] M W Layton, S A Goldstein, R W Goulet, L A Feldkamp, D J Kubinski, and G G Bole. Examination of subchondral bone architecture in experimental osteoarthritis by microscopic computed axial tomography. *Arthritis Rheum*, 31(11):1400–5, 1988.

[93] Ryan Lee and Walter Kean. Obesity and knee osteoarthritis. *Inflammopharmacology*, 20(2):53 – 58, 2012.

[94] Hongdong Li, Yizeng Liang, Qingsong Xu, and Dongsheng Cao. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Analytica Chimica Acta*, 648(1):77–84, 2009.

[95] Elia Liitiainen, Francesco Corona, and Amaury Lendasse. On the curse of dimensionality in supervised learning of smooth regression functions. *Neural Processing Letters*, 34:133–154, 2011.

[96] J C Lin, M Amling, D C Newitt, K Selby, S K Srivastav, G Delling, H K Genant, and S Majumdar. Heterogeneity of trabecular bone structure in the calcaneus using magnetic resonance imaging. *Osteoporos Int*, 8(1):16–24, 1998.

[97] Huan Liu and Hiroshi Motoda. *Feature Extraction, Construction and Selection: A Data Mining Perspective*. Kluwer Academic Publishers, Norwell, MA, USA, 1998.

[98] M. Loog and B. Van Ginneken. Static posterior probability fusion for signal detection: applications in the detection of interstitial diseases in chest

radiographs. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 1, pages 644–647, 2004.

[99] John A. Lynch, Souhil Zaim, Jenny Zhao, Charles G. Peterfy, and Harry K. Genant. Automating measurement of subtle changes in articular cartilage from mri of the knee by combining 3d image registration and segmentation. *Proceedings of SPIE, Medical Imaging*, pages 431–439, 2001.

[100] S Majumdar, D Newitt, M Jergas, A Gies, E Chiu, D Osman, J Keltner, J Keyak, and H Genant. Evaluation of technical factors affecting the quantification of trabecular bone structure using magnetic resonance imaging. *Bone*, 17(4):417–30, 1995.

[101] Oded Maron and Tomás Lozano-Pérez. A framework for multiple-instance learning. In *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, pages 570–576. MIT Press, 1998.

[102] Joselene Marques, Harry K. Genant, Martin Lillholm, and Erik B. Dam. Diagnosis of osteoarthritis and prognosis of tibial cartilage loss by quantification of tibia trabecular bone from mri. *Magnetic Resonance in Medicine*, 2012.

[103] Joselene Marques, Rabia Granlund, Martin Lillholm, Paola C. Pettersen, and Erik B. Dam. Automatic analysis of trabecular bone structure from knee mri. *Computers in Biology and Medicine*, (0):–, 2012.

[104] Tahir Mehmood, Harald Martens, Solve Saebo, Jonas Warringer, and Lars Snipen. A partial least squares based algorithm for parsimonious variable selection. *Algorithms for Molecular Biology*, 6(1):27, 2011.

[105] R. W. Moskowitz, R. D. Altman, M. C. Hochberg, J. A. Buckwalter, and V. M. Goldberg. *Osteoarthritis: Diagnosis and Medical/Surgical management*. Wolters Kluver Health, fourth edition, 2007.

[106] R.W. Moskowitz, R.D. Altman, J.A. Buckwalter, V.M. Goldberg, and M.C. Hochberg. *Osteoarthritis: Diagnosis and Medical/Surgical Management*. Doody's all reviewed collection. Lippincott Williams & Wilkins, fourth edition, 12 2006.

[107] Lopez AD Murray CJL. The global burden of disease. In *World Health Organization*, 1997.

[108] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, October 2010.

[109] A. M. Parfitt, M. K. Drezner, F. H. Glorieux, J. A. Kanis, H. Malluche, P. J. Meunier, S. M. Ott, and R. R. Recker. Bone histomorphometry: standardization of nomenclature, symbols and units. *Journal of Bone Mineral Research*, 2(6):595–610, 1987.

[110] Jean-Pierre Pelletier, Jean-Pierre Raynauld, Marie-Josee Berthiaume, Francois Abram, Denis Choquette, Boulos Haraoui, John Beary, Gary Cline, Joan Meyer, and Johanne Martel-Pelletier. Risk factors associated with the loss of cartilage volume on weight-bearing areas in knee osteoarthritis patients assessed by quantitative magnetic resonance imaging: a longitudinal study. *Arthritis Research & Therapy*, 9(4):R74, 2007.

[111] C. G. Peterfy, A. Guermazi, S. Zaim, P. F. Tirman, Y. Miaux, D. White, M. Kothari, Y. Lu, K. Fye, S. Zhao, and H. K. Genant. Whole-organ magnetic resonance imaging score (WORMS) of the knee in osteoarthritis. *Osteoarthritis and Cartilage*, 12(3):177–190, 2004.

[112] StephenM. Pizer, P.Thomas Fletcher, Sarang Joshi, Andrew Thall, JamesZ. Chen, Yonatan Fridman, DanielS. Fritsch, A.Graham Gash, JohnM. Glotzer, MichaelR. Jiroutek, Conglin Lu, KeithE. Muller, Gregg Tracton, Paul Yushkevich, and EdwardL. Chaney. Deformable m-reps for 3d medical image segmentation. *International Journal of Computer Vision*, 55:85–106, 2003.

[113] P. Pudil, J. Novovičová, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11):1119–1125, 1994.

[114] A. A. Qazi, J. Folkesson, P. C. Pettersen, M. A. Karsdal, C. Christiansen, and E.B. Dam. Separation of healthy and early osteoarthritis by automatic quantification of cartilage homogeneity. *Osteoarthritis and Cartilage*, 15(10):1199–1206, 2007.

[115] Arish A. Qazi, Erik B. Dam, Mads Nielsen, Morten A. Karsdal, Paola C. Pettersen, and Claus Christiansen. Osteoarthritic cartilage is more homogeneous than healthy cartilage: Identification of a superior region of interest colocalized with a major risk factor for osteoarthritis. *Academic Radiology*, 14(10):1209 – 1220, 2007.

[116] P. Qvist, A. C. Bay-Jensen, C. Christiansen, E. B. Dam, P. Pastoureau, and M. A. Karsdal. The disease modifying osteoarthritis drug (DMOAD): Is it in the horizon? *Pharmacological Research*, 58:1–7, 2008.

[117] E L Radin, I L Paul, and M J Tolkoff. Subchondral bone changes in patients with early degenerative joint disease. *Arthritis Rheum*, 13(4):400–5, 2005.

[118] E L Radin and R M Rose. Role of subchondral bone in the initiation and progression of cartilage damage. *Clin Orthop Relat Res*, (213):34–40, 1986.

[119] J. Ramirez, J.M. Gorriz, F. Segovia, R. Chaves, D. Salas-Gonzalez, M. Lopez, I. Alvarez, and P. Padilla. Computer aided diagnosis system for the alzheimer's disease based on partial least squares and random forest spect image classification. *Neuroscience Letters*, 472(2):99 – 103, 2010.

[120] J. Raundahl, M. Loog, P. Pettersen, L.B. Tanko, and M. Nielsen. Automated effect-specific mammographic pattern measures. *Medical Imaging, IEEE Transactions on*, 27(8):1054–1060, 2008.

[121] J. Raundahl, M. Loog, P. C. Pettersen, L. B. Tanko, and M. Nielsen. Automated effect-specific mammographic pattern measures. *IEEE Transactions on Medical Imaging*, 27(8):1054–1060, 2008.

[122] J. Raundahl, Marco Loog, P. Pettersen, L.B. Tank, and Mads Nielsen. Automated effect-specific mammographic pattern measures. *IEEE Transactions on Medical Imaging*, 27(8):1054–1060, 2008.

[123] M. Reijman, J. M. W. Hazes, S. M. A. Bierma-Zeinstra, B. W. Koes, S. Christgau, C. Christiansen, A. G. Uitterlinden, and H. A. P. Pols. A new marker for osteoarthritis: cross-sectional and longitudinal approach. *Arthritis & Rheumatism*, 50(8):2471–2478, 2004.

[124] Joseph Lee Rodgers and W Alan Nicewander. Thirteen ways to look at the correlation coefficient. *American Statistician*, 42(1):59, 1988.

[125] F. W. Roemer and A. Guermazi. MR imaging-based semiquantitative assessment in osteoarthritis. *Radiologic Clinics of North America*, 47(4):633–654, July 2009.

[126] F W Roemer, A Guermazi, M K Javaid, J A Lynch, J Niu, Y Zhang, D T Felson, C E Lewis, J Torner, and M C Nevitt. Change in mri-detected subchondral bone marrow lesions is associated with cartilage loss: the most study. a longitudinal multicentre study of knee osteoarthritis. *Ann Rheum Dis*, 68(9):1461–5, 2009.

[127] Hanna Fahad S, Teichtahl Andrew J, Wluka Anita E, Wang Yuanyuan, Urquhart Donna M., English Dallas R, Giles Graham G., and Cicuttini Flavia

M. Women have increased rates of cartilage loss and progression of cartilage defects at the knee than men: a gender study of adults without clinical knee osteoarthritis. *Menopause*, 19:666?670, 2009.

[128] Jonathan R Shewchuk. An introduction to the conjugate gradient method without the agonizing pain. Technical report, Pittsburgh, PA, USA, 1994.

[129] Stuart Solloway, Charles E. Hutchinson, John C. Waterton, and Christopher J. Taylor. The use of active shape models for making thickness measurements of articular cartilage from mr images. *Magnetic Resonance in Medicine*, 37(6):943–952, 1997.

[130] P. Somol, P. Pudil, J. Novovičová, and P. Paclík. Adaptive floating search methods in feature selection. *Pattern Recognition Letters*, 20(11-13):1157–1163, 1999.

[131] M. Sonka and J. M. Fitzpatrick. *Handbook of Medical Imaging - Medical Image Processing and Analysis*, volume 2. SPIE, 2000.

[132] Lars Sørensen, Shaker S. B., and Marleen de Bruijne. Quantitative analysis of pulmonary emphysema using local binary patterns. *IEEE Transactions on Medical Imaging*, 29(2):559–569, 2010.

[133] Kenji Suzuki, Pingkun Yan, Fei Wang, and Dinggang Shen. Machine learning in medical imaging. *Int. J. Biomedical Imaging*, pages –1–1, 2012.

[134] Henrik S. Thomsen, Lone Larsen, Elizaveta Chabanova, and Jakob M. Moller. Open low-field-strength mri of the shoulder is not so bad. *American Journal of Roentgenology*, 182(6):1601–1602, 2004.

[135] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994.

[136] B. Vasilic and F.W. Wehrli. A novel local thresholding algorithm for trabecular bone volume fraction mapping in the limited spatial resolution regime of in vivo mri. *Medical Imaging, IEEE Transactions on*, 24(12):1574 –1585, 12 2005.

[137] Xiaogang Wang and Xiaoou Tang. Experimental study on multiple lda classifier combination for high dimensional data classification.

[138] Joachim Weickert. *Anisotropic Diffusion in Image Processing*. B.G.Teubner Stuttgart, 1998.

[139] Miles Wernick, Yongyi Yang, Jovan Brankov, Grigori Yourganov, and Stephen Strother. Machine Learning in Medical Imaging. *IEEE Signal Processing Magazine*, 27(4):25–38, 2010.

[140] Lukas Martin Wildi, Jean-Pierre Raynauld, Johanne Martel-Pelletier, André; Beaulieu, Louis Bessette, Fréderic Morin, Francois Abram, Marc Dorais, and Jean-Pierre Pelletier. Chondroitin sulphate reduces both cartilage volume loss and bone marrow lesions in knee osteoarthritis patients starting as early as 6 months after initiation of therapy: a randomised, double-blind, placebo-controlled pilot study using mri. *Ann Rheum Dis*, 70(6):982–9, 2011.

[141] Svante Wold, Michael Sjostrom, and Lennart Eriksson. Pls-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109 – 130, 2001.

[142] Svante Wold and Johan Trygg. The pls method – partial least squares projections to latent structures – and its applications in industrial rdp (research , development , and production). *PLS in industrial RPD for Prague*, 1(June):1–44, 2004.

[143] Yunyan Zhang. Mri texture analysis in multiple sclerosis. *Journal of Biomedical Imaging*, 2012:2:2–2:2, 2012.

List of Publications

Papers in international journals

- J. Marques, C. Igel, M. Lillholm, E. B. Dam, "Linear feature selection in texture analysis - A PLS based method," Machine Vision and Applications, Special Issue on Machine Learning in Medical Imaging, 2012, DOI: 10.1007/s00138-012-0461-1

- J. Marques, Harry K. Genant, M. Lillholm, E. B. Dam, "Diagnosis of osteoarthritis and prognosis of tibial cartilage loss by quantification of trabecular Tibia bone from MRI," Magnetic Resonance in Medicine, 2012, DOI: 10.1002/mrm.24477

- J. Marques, R. B. Granlund, M. Lillholm, P. C. Pettersen, E. B. Dam, "Automatic Analysis of Trabecular Bone Structure from Knee MRI," Computers in Biology and Medicine, Vol. 42, Issue 7, Pages 735-742, DOI: 10.1016/j.compbiomed.2012.04.005, July 2012.

- J. Marques, D. M. J. Tax, M. Loog, E. Dam, "The most informative texture region of the tibia for predicting cartilage loss - A multiple-instance learning approach", Journal of IEEE Transaction on Medical Imaging (submitted)

Papers in conference proceedings

- J. Marques, L. Clemmensen, E. B. Dam, "Diagnosis and prognosis of osteoarthritis by texture analysis using sparse linear models", Proceedings of the Workshop on Sparsity Techniques in Medical Imaging (STMI) in conjunction with 15th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI).

- J. Marques, E. B. Dam, "Texture analysis by a PLS based method for combined feature extraction and selection" Proceedings of the 2nd Workshop on Machine Learning in Medical Imaging (MLMI) in conjunction with 14th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Lecture Notes in Computer Science. Toronto, Canada, Sep 2011. p. 109-116.

Published abstracts

- E. B. Dam, J. Marques, S. Zaim, T. Fuerst, H. Genant, M. Lillholm, M. Nielsen, "Fully automatic cartilage morphometry for knee MRI from the OAI", Workshop on Osteoarthritis Imaging. Hilton Head, South Carolina, Jul 2012.

- E. B. Dam, J. Marques "Automatic segmentation of bone and cartilage from knee MRI", 97th Scientific Assembly and Annual Meeting Radiological Society of North America (RSNA), Chicago, Illinois, Nov 2011.

- J. Marques, E. B. Dam, "Prognosis of cartilage loss by analyzing magnetic resonance imaging of the Tibia trabecular bone structure", World Congress on Osteoarthritis (OARSI). San Diego, California, Sep 2011. p. S9(7)

- J. Marques, E. B. Dam, "Diagnosis of osteoarthritis by quantification of Tibia trabecular bone structure from magnetic resonance imaging", World Congress on Osteoarthritis (OARSI). San Diego, California, Sep 2011. p. S75(149)

- J. Marques, E. B. Dam, "Prediction of cartilage loss by analysis of trabecular Tibia bone structure", Workshop on Osteoarthritis Imaging. Salzburg, Austria, Jun 2011, p.42

- E. B. Dam, J. Marques "Automatic segmentation of bone and cartilage from knee MRI", Workshop on Osteoarthritis Imaging. Salzburg, Austria, Jun 2011, p.55

- J. Marques, E. B. Dam, "Quantification of trabecular Tibia bone structure related to the presence of Osteoarthritis", Workshop on Osteoarthritis Imaging. Salzburg, Austria, Jun 2011, p.60

Acknowledgements