Predictive Modelling of Femur Fracture from DXA Scans Using Radiomics and Machine Learning

Raquel Rodríguez Martín
Predictive modelling of femur fracture from DXA scans using radiomics and machine learning

Raquel Rodríguez Martín

BACHELOR THESIS UPF / 2018

THESIS SUPERVISOR

Dr. Karim Lekadir

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGIES
Acknowledgments

I would like to thank everyone who helped in the completion of this project either by direct counseling, material facilitation or emotional support.
Summary

Predictive modeling of bone fractures at the hip using DXA images is an important research field due to the capability of this imaging modality to capture osteoporosis related changes in the bone tissues. There exist many available techniques providing fracture risk assessment, such as bone mineral density measurements and biomechanical models. However, these methods do not use the wealth of information provided by the DXA scans and thus lack the accuracy to enable their translation to clinical practice.

In this study, a radiomics and machine learning approach is proposed for a more comprehensive predictive modelling of femur fracture using DXA. Our main hypothesis is that integrating heterogeneous and complex characteristics of the bone tissue through radiomics at both the global and local scales will lead to improved prediction of fracture risk. In the proposed technique, the optimal radiomics indices of different types (shape, intensity and texture based) are selected using feature selection methods to identify the most relevant ones for discriminating low-risk and high-risk cases. Furthermore, advanced machine learning is applied to integrate the selected radiomics features into a unified risk classification model based on different learning models (Support Vector Machines, Decision Tree and Random Forest).

The proposed predictive model was validated using 63 cases including to patients with and without femur fracture. In this preliminary study, all cases were correctly classified using the proposed model, indicating great potential of radiomics-based classification for predicting fractures of the femur.

Keywords

Radiomics features, classification model, hip fracture
Index

1. INTRODUCTION .................................................................................................................. 1
   1.1. Osteoporosis Disease ................................................................................................. 1
   1.2. Prediction of bone fracture ....................................................................................... 2
   1.2.1. Dual X-ray absorptiometry and BMD estimation .................................................. 2
   1.2.2. Biomechanical indices of skeletal metabolism ..................................................... 4
   1.2.3. Clinical risk factors ............................................................................................... 4
   1.2.4. Biomechanical models ......................................................................................... 5
   1.3. Proposed Approach .................................................................................................. 6
   1.3.1. Radiomics ........................................................................................................... 6
   1.4. Motivation and Thesis Objectives ............................................................................ 9

2. METHODOLOGY .................................................................................................................. 10
   2.1. Methodology Overview ............................................................................................ 10
   2.2. Data Description ....................................................................................................... 11
   2.3. Data Processing ......................................................................................................... 11
   2.3.1. Segmentation and mask generation ....................................................................... 11
   2.3.2. Registration process ............................................................................................... 12
   2.4. Input of the predictive model of fracture risk .......................................................... 14
   2.4.1. Radiomics features for fracture prediction ........................................................... 15
   2.4.2. Clinical variables .................................................................................................. 16
   2.5. Construction of the predictive model ......................................................................... 16
   2.5.1. Normalization of the input of the predictive model .............................................. 17
   2.5.2. Machine learning models ...................................................................................... 17
   2.5.3. Feature Selection ................................................................................................ 19
   2.6. Model validation ......................................................................................................... 19

3. RESULTS AND DISCUSSION .............................................................................................. 20
   3.1. Evaluation of the registration process ....................................................................... 20
   3.2. Classification results per bone structure .................................................................... 20
   3.3. Results per region within the femur ......................................................................... 21
   3.4. Impact of the number of radiomic features ............................................................... 22
   3.5. Incorporating clinical variables to the predictions .................................................... 23
   3.6. Analysis of radiomics types ...................................................................................... 23
   3.7. Impact of the feature selector .................................................................................... 24
   3.8. Best Selected Features ............................................................................................. 25

4. CONCLUSIONS .................................................................................................................... 27

5. BIBLIOGRAPHY .................................................................................................................. 29
List of Figures

Figure 1. Pipeline of radiomics process, including image acquisition and segmentation, extraction of quantitative image biomarkers, advanced machine learning and application to personalized medicine [1] .......................... 7

Figure 2. Overview of the six main steps involved in the proposed approach ............................................................................................................................................................................. 10

Figure 3. a) Original data provided by CETIR, automatic segmentation approach contouring the femur and pelvis bone. a) Correction of the original segmentation approach delineating femur and pelvis bone. a) Segmentation of the femur bone .......................................................................................................................... 11

Figure 4. a) DXA of the segmented femur visualization with ITK Snap tool. b) DXA of the segmented femur with its corresponding mask visualization using ITK Snap tool .................................................................................................................................................. 12

Figure 5. a) DXA segmentation of the femur, corresponding to the fixed image in the registration process. b) Image containing regions of major interest, corresponding to the moving image in the registration process............ 13

Figure 6. a) Mask visualization corresponding to the registration of the femoral neck region (showed in color red using ITK Snap tool). b) Definition of the five regions to be registered ........................................................................................................................................ 14

Figure 7. Accuracy of the SVM classification model as a function of the number of features selected in the model ................................................................................................................................................................. 22

Figure 8. Classification accuracy using SBFS as a function of 525 features .................................................................................................................................................................................................................. 25

Figure 9. New regions of interest for radiomics-based prediction of bone fracture ........................................................................................................................................................................................................ 28
List of Tables

Table 1. List of clinical variables tested within the classification model................................................................................................................................. 16

Table 2. Classification accuracy for features extracted from different bone structures (femur and pelvis bones).................................................................................................................. 20

Table 3. Classification accuracy by using feature combinations extracted from different regions.................................................................................................................. 21

Table 4. Classification accuracy including clinical variables to the input of the model.......................................................................................................................... 23

Table 5. Classification accuracy of features from different classes............................................................................................................................... 23

Table 6. Analysis of the 30 best selected features and its relevance................................................................................................................................. 25
1. INTRODUCTION

1.1. Osteoporosis Disease

Osteoporosis is a skeletal disease that affects the microarchitecture of bone tissue. Patients with osteoporosis suffer either from bone tissue deterioration or low bone mass density, resulting in bones that are more porous, weaker and susceptible to fracture. These changes can be assessed by computing the difference of the bone mass at any time-point of the adulthood to the one present at the maturation stage [2]. As bone mass decreases over time, the prevalence of osteoporosis rises with aging populations. Moreover, due to the higher life expectancy, as well as the lack of estrogen after menopause as consequence of hormone alterations, this systemic disease is three times more prevalent amongst women [3].

Osteoporosis is estimated to affect approximately around 75 million people between Europe, USA and Japan [4]; causing around 8.9 million fractures worldwide [5], including 1/3 in women older than 50 and 1/5 in men of the same condition [6], [7]. The main clinical challenge with osteoporosis is the lack of clinical evidence or symptoms until a fracture occurs, leading to missing diagnosis in many cases. These fractures have a negative impact on morbidity, reducing quality of life and increasing mortality rate due to associated complications depending on each patient’s condition [8], [9]. Furthermore, osteoporosis associated fractures not only involve difficulties for the patient but also imply direct health care costs of approximately €32 billion in the EU and it is expected that the expenditure rate doubles its value by year 2050 [10].

The worse consequence of osteoporosis is hip fracture, as it is most painful and disabling over long periods of time. Almost in every case there is a need for hospitalization with a mean stay of 30 days [11]. Even though almost all fractures of this kind heal well, immobility can lead to a higher risk of complications and long-term care services. A study showed that after six months of suffering fracture, 85% of patients could not walk unaided and 24% of the patients older than 50 died in the following year [12]. The lifetime risk of suffering a hip fracture lies between 14% and 20% for Caucasian women of Europe and USA [13]. From the 8.9 million fractures suffered by the world population mentioned above, 1.6 million were at the hip [5], and this rate is expected to rise in the following years.
Nowadays there exist guidelines to prevent, diagnose and treat osteoporosis, although there is not a standard procedure accepted worldwide. Both the International Committee for Osteoporosis Clinical Guidelines [14] and WHO guidelines for osteoporosis [15] define procedures and suggestions to identify patients suffering from osteoporosis, treatment specially for postmenopausal women and routine screening. These guidelines are based on an analysis of risk factors, although in many cases the disease is not diagnosed even after a fracture occurred or patients do not receive treatment even though an existing osteoporosis is diagnosed.

Given the major consequences of osteoporosis as explained above, an early diagnosis would enable timely intervention by clinicians to reduce the probability of suffering a fracture and a decline in life quality. Furthermore, not only it would be advantageous for patients but for the cost-effectiveness of the health care system. This thesis will use bone imaging data and machine learning, as well as advanced imaging phenotyping, to develop a new accurate technique for fracture risk assessment.

1.2. Prediction of bone fracture

The main techniques used by clinicians and researchers to identify patients at risk of suffering a hip fracture, and thus to enable fracture prevention, are discussed as follows.

1.2.1. Dual X-ray absorptiometry and BMD estimation

Nowadays the most common way to predict fractures in clinical practice is by measuring the Bone Mass Density (BMD) based on a dual x-ray absorptiometry (DXA) scan with the aim of anticipating the fracture risk as consequence of osteoporosis. DXA is considered the “gold standard” method to assess osteoporosis status and evaluate the fracture risk, although other techniques as ultrasound, radiography and computed tomography have been also studied. Note that these imaging techniques give results with lower sensitivity compared to DXA, although with CT is a valuable approach the patient receives more radiation than when using DXA [16].

A DXA scan consists of an open X-ray table where the patient lies and two X-ray beams with different energy levels are focused on the bones of interest. The machine measures the amount of energy that passes through the bone from both beams (high and low energy). Afterwards, the bone mineral content can be measured based on the energy differences obtained between the two beams. The acquired values are directly dependent
of the size and integral density of the bone tissue, so BMD is derived from dividing the bone mineral content by the area or volume measured.

The BMD measurement can be interpreted with the definition of osteoporosis proposed by the WHO, using the so-called T-score threshold [15], which indicates the relative density compared to that of a healthy 30-year-old when bones are at their strongest. Patients can be classified by using the results of the T-score: the WHO established that a T-score of -1.0 or higher stands for a healthy bone, while a T-score between -1.0 and -2.5 is diagnosed as osteopenia and a value below -2.5 corresponds to osteoporosis disease.

In clinical practice, current osteoporosis diagnosis is mainly based on BMD measurements as several prospective studies have shown that there is a significant correlation between low BMD and the frequency of suffering a fracture [17]. It has been established in many prospective studies that measurements of T-score at the hip are the best predictors of hip fracture, thus providing the best diagnostic criteria [18]. However, due to the variability amongst the population and the distinct rates of bone loss depending on the measured site, T-score values cannot be easily interpreted at different bone sites [19]. Moreover, relating correlations of measures in different sites should be avoided for predictive purposes as this has been shown to lead to miss-classifications [20].

It is important to note that BMD has important limitations as it does not consider other factors including skeletal abnormalities and the geometry of the bone which are essential to obtain a better estimation of the fracture risk [21]. BMD techniques based on absorptiometry are highly specific but less sensitive, as it can vary depending on the definition of the risk thresholds to use. Many problems could arise if the fracture risk thresholds are influenced by the reimbursement of costs by health care agencies whenever a DXA scan is performed. In other words, some established thresholds for treatment of osteoporosis in many countries could not be fully reliable [3].

The performance of BMD in predicting fracture is, however, at least as good as that of blood pressure in predicting stroke, and considerably better than the use of serum cholesterol to predict coronary artery disease [15]. However, many fractures occur in patients that have not been diagnosed with osteoporosis, meaning that a normal BMD value does not guarantee that a fracture will not occur. While BMD measurements by DXA scans improve the prediction of fractures; it does not perform so well for some cases such as postmenopausal women [3]. There is also great concern expressing risk
assessment in a standardized way and in absolute terms to avoid confusion in clinical practice.

1.2.2. **Biomechanical indices of skeletal metabolism**

Biochemical indices of skeletal metabolism are of clinical relevance in assessing risk of bone fracture as indicative bone turnover. For example, biochemical markers have potential clinical value for the estimation of fracture probability in women as they experience hormone alterations after menopause. Furthermore, these biochemical markers are easier and cheaper to obtain in the current clinical practice when compared to densitometry techniques. Some studies show that bone loss is correlated to some biochemical indices of bone resorption, although these changes are not sufficient to discriminate patients with osteoporosis [22]. Therefore, it is highly probable that the combination of these markers with BMD measurements is likely to improve the prediction accuracy for osteoporotic fractures.

There are biochemical markers of bone turnover of two different kinds: bone resorption markers and bone formation. The most common bone resorption markers are hydroxyproline, pyridinium cross-links, and their peptides, while on the bone formation marker side the most used are total alkaline phosphatase, the bone isoenzyme of alkaline phosphatase, osteocalcin, and the procollagen propeptides of type I collagen. Some studies have evaluated biochemical bone turnover indices for their predictability on their own and by combining them together with BMD measurements [22], [23].

There are some limitations in fracture predictions based on biochemical markers are they are influenced by food intake. In addition, bone resorption markers vary in consonance with the circadian rhythm, while bone formation markers are less susceptible [24]. This biological variation can make it difficult to interpret the assessment of fracture risk. Another drawback of using biochemical methods is that bone markers can be altered by previous fractures. While some studies show that indices of bone metabolism can predict clinical symptomatic vertebral fractures, it has also been seen that they have no predictability in the hip. This is probably because of the higher metabolism activity in the spine compared with the hip [22].

1.2.3. **Clinical risk factors**

Many clinical factors have been used to assess fracture risk, including age, gender, early menopause, previous fractures, family history, low body weight, specific diseases such
as hypogonadism, use of corticosteroids, conditions increasing the risk of fall, smoking and physical inactivity. There are guidelines established by the International Committee for Osteoporosis that analyze clinical risk factors as to detect potential osteoporotic individuals, albeit many cases are not diagnosed on time. For a better management of the disease, some tools that combined clinical risk factors with BMD measurements have been proposed. An example is the algorithm developed by the WHO for intervention thresholds for osteoporosis in the clinical routine [25]. Another example is the FRAX tool which combines the BMD measure at femur neck with models of patient from different cohorts that integrate risk factors with the aim of enhancing the accuracy of 10-year hip fracture prediction [26].

Age has been reported to be the most powerful independent risk factor [12]. On the other side, clinical risk factors are relatively poor when predicting BMD and fracture risk by themselves [3]. Moreover, risk factors have different impacts depending on the type of fracture, age group, ethnicity and geographic regions of the evaluated patients. This fact suggests that to accurately predict future fracture, these need to be validated on large populations, adjusted for gender, age and the type of fracture. Finally, it is important to consider that the relation between some of the risk factors has not been fully studied nor validated, which means there is limited knowledge of the impact they have on predicting the risk of fracture [27].

### 1.2.4. Biomechanical models

Up to this point the different techniques described in this thesis for fracture assessment consider bone-related parameters together with risk factors in some cases. Generally, BMD is calculated from medical images, however this measurement cannot indicate any other mechanical property than bone strength [21]. Biomechanical models add precision to the assessment of fracture risk by adding patient-specific parameters as quality of the bone, proximal femur anatomy, cortical thickness, geometry, loading conditions, etc. It has been shown that these specific factors affect the fracture of the hip [28]. However, there are many biomechanical models that do not consider the parameters mentioned above as this implies adding levels of complexity to the model. There is a lack of models that comprise bone-related parameters, risk factors as well as fall and loading conditions to accurately predict hip fracture.

Personalized biomechanical models are built from medical images, mainly DXA and CT. Sarvi et al. [28] proposed a precise, fast and comprehensive biomechanical model created
from DXA scans for predicting fracture risk at hip. The model consisted of two biological levels for the whole body and the proximal femur, while all available parameters regarding the patient were added to complete the model. Side-fall simulations were afterwards performed. The model proposed by López et al. [29] simulated bone generation and changes due to aging for predicting bone turnover and BMD loss by using a finite element (FE) method. However, such biomechanical models have their own drawbacks. In general, they are time consuming (they can run up-to a few hours/days depending on the complexity and resolution), while predictability of FE analysis differs a lot by using different mesh elements size [30]. In addition, FE analysis models perform better for predictions with elderly patients at risk of suffering a fracture in the proximal femur [31]. Finally, this type of techniques compute prediction for hip fractures only if a fall occurs.

1.3. Proposed Approach

Several techniques for fracture prediction used by practitioners and researchers have been discussed above. Notwithstanding, it has been observed that these methods have limitations which reduce their prediction accuracy. Each technique uses specific parameters to make predictions, such as BMD, biochemical markers, or biomechanical indices, while in some cases combinations are made. In particular, the use of medical image data through DXA is promising as it allows direct measurement of the bones, but thus far the existing techniques have not exploited all the wealth of information that is available in DXA. For example, BMD which is widely used in clinical practice provide a global measure of the bone strength and thus an incomplete view of the bone’s state, ignoring important information on the local geometric and tissue characteristics of the bones. Our hypothesis in this thesis is that using a more advanced image analysis approach to quantify additional characteristics of the bone from DXA at both the global and local scales will lead to an improved prediction of fracture risk. To achieve this, we propose the very first radiomics and machine learning approach for a more comprehensive prediction of femur fracture using DXA.

1.3.1. Radiomics

Lambin et al. define radiomics as “the high-throughput extraction of large amounts of image features from radiographic images” [32]. The suffix –omics is widely used in medical research fields and it is referred to the generation of high-dimensional data from
single samples [33]. The hypothesis of radiomics relies on the fact that images are more than data and can serve not only for visualization but for extracting mineable data describing pathophysiology [1]. Specifically, radiomics correspond to a large pool of quantitative imaging features that describe a wide range of tissue properties, including shape, size, boundary, intensity and texture-based. Radiomics is an innovative way of analyzing medical images, leading to a comprehensive quantification and deeper phenotyping of disease. Up to the moment, this approach has been mainly addressed to oncology problems with the objective of improve diagnose of cancer and monitor response to treatment [1], [32], [34].

The workflow of radiomics includes several steps (Fig. 1). Firstly, it requires the acquisition of a high-quality medical image from any modality following optimum protocols. Then an experienced physician localizes the region or volume of interest and segments it or supervises an automated segmentation. The process of delineating the structures of interest is as challenging as critical, because will determine the regions at which the radiomic indices are extracted. However, anatomical structures and medical image quality differ a lot between patients [1]. Afterwards, radiomics image features describing signatures from the defined region of interest are extracted. Later, these indices are selected using machine learning according to their relevance to the clinical problem being evaluated. Those most informative selected features are analyzed to understand their relevance to diagnosis, prediction or prognosis. Finally, the obtained rediomics can be combined using machine learning to obtain a model of risk of a certain disease, i.e. by

![Fig. 1. Pipeline of radiomics process, including image acquisition and segmentation, extraction of quantitative image biomarkers, advanced machine learning and application to personalized medicine [1].](image-url)
developing a classifier able to make prediction on its own or as part of an algorithm integrating other types of data (e.g. clinical data, risk factors, etc).

Gillies et al. defined the features extracted in the radiomics process as “semantic” or “agnostic”. The semantic ones describe regions of interest and are used by radiologists, some examples are size and shape. While the agnostic indices quantitatively describe tissue heterogeneity, are mathematically obtained and are not part of radiologists’ lexicon, as examples we have histogram based features and wavelets [1]. In currently clinical routine, images are used as visualization tools to detect features of interest, which means the lack of a standard acquisition and reconstruction does not interfere in the task of practitioners. However, when extracting radiomic information as indices of pathological variation, a protocol must be defined to ensure that the extracted features describe tissue variation instead of acquisition alterations.

In this thesis, radiomics is intended to be used as a decision support tool by its own or complementing other models, thereby any available data from patients can be expressed as a radiomic feature (demographic, clinical, comorbidity or genomic data); in this manner diagnosis and prognosis can be potentially improved by considering such a large amount of information [35]. Building large databases is of maximum importance, as the performance of the predictive model is dependent of the size and quality of the dataset. An advantage of using radiomics as a standard-of-care imaging technique is its need of images that together with the pool of features extracted could build a big data-base source for further interpretations and study contributions. Moreover, detection of patterns within diseases could be achieved with big data sources.

Radiomics field has been widely explored in cancer imaging quantification, specifically for treatment response and for estimation of patient prognosis [36]. Radiomics applications have been mainly focused on oncology because their study can serve to understand the existing relationships between imaging features of tumors. Aerts et al. [36] studied a radiomics approach in oncology consisting on evaluating the value of radiomics indices to differentiate phenotypic differences of tumors. Their results showed that a large number of radiomics features have prognostic potential for independent data sets of lung and head-and-neck cancer screening.

Radiomics have also been used in non-cancer disease approaches. Oakden-Rayner et al. [37] studied longevity predictability using feature engineering and deep learning
methods. They hypothesized that underlying microscopic changes occurring in body tissue could be quantified by using a high-throughput extraction of radiomics indices. Finally, Cetin et al. [38] presented a new radiomics approach to identify cardiovascular diseases from cine-MRI and thus enable improved stratification of CVD patients. They calculated large pools of radiomics features encoding relevant changes in anatomical and image characteristics of the ventricles due to CVDs. Sequential forward feature selection was used to identify the most relevant descriptors for given CVD classes. Finally, advanced machine learning is applied to suitably integrate the selected radiomics for final multi-feature classification. They achieved 92% classification accuracy with their method.

1.4. Motivation and Thesis Objectives

The work presented in this thesis is motivated by the need for more accurate and comprehensive methods to enrich conventional methods in the osteoporosis disease assessment field. With the proposed radiomics approach, information usage is maximized and additional knowledge is generated from feature extraction of DXA images. As mentioned before, radiomics has mainly been applied to oncology problems and this thesis is the first attempt to apply the paradigm to DXA images with the aim of predicting osteoporosis related fractures. The main goals of the project are:

1. Performing radiomics analysis to extract subtle and advanced imaging signatures from DXA scans of the femur and to explore potential applications for osteoporotic fracture prediction. Our hypothesis is that bone changes due to osteoporosis is reflected in absorptiometric images that can be quantitatively obtained by extracting advanced image features based on radiomics.

2. Building a classification model that combined the extracted radiomics features with the aim of aiding the diagnose of bone disease and decision support of risk prediction. In addition, we will explore its potential predictability and fracture risk assessment by further analyzing the most relevant selected features.

3. Analyzing the performance of the predictive model when adding clinical and patient-specific variables. Our hypothesis was that the combination of radiomics with clinical data such as age, gender, and lifestyle, could further improve decision support on risk fracture assessment.
A complete understanding of radiomics procedure together with an analysis of extracted features, may ultimately help clinicians to predict risk fracture on time by integrating richer imaging information to their routine clinical practice.

The project hereby presented emerges from the Departament de Tecnologies de la Informació i les Comunicacions (DTIC) from Pompeu Fabra University, Barcelona, Spain, in collaboration with Dr. Luis Miguel del Rio Barquero from CETIR medical center also located in Barcelona, Spain.

2. METHODOLOGY

2.1. Methodology Overview

The proposed approach is divided into six steps with the goal of building a predictive model for osteoporosis related fracture risk based on radiomics (see Fig. 2), i.e.

1. Image data collection (Section 2.2).
2. Segmentation of the femur (Section 2.3.1).
3. Image registration to define the regions within the femur for estimating the radiomics (Section 2.3.2).
4. Extraction and selection of the radiomics variables (Section 2.4.1)
5. Construction of the predictive model using machine learning (Section 2.5)
6. Model validation (Section 3).

Fig. 2. Overview of the six main steps involved in the proposed approach.
2.2. Data Description

This study was conducted with data set collected by Dr. Luis Miguel del Río Barquero at CETIR medical centre, Barcelona, Spain. The database consists of DXA scans from 65 patients, including 30 patients who suffered a hip fracture (either at femoral neck or at trochanter; see Table 1) while the remaining 35 were acquired from patients who did not experience a fracture. Those scans corresponding to patients who sustained a fracture were taken before the fracture occurred. Two of the images were excluded from the study because they presented artifacts on the image during the pre-processing achieved by the clinicians. The final dataset used in the study was of 63 DXA scans in TIFF image type format (30 for fracture and 33 for non-fracture).

Together with the images, a list of relevant clinical data was also provided, including sex, age, weight, height, hip tissue thickness, fragility fractures, family history fractures, other diseases suffered by the patients and the mean spine trabecular bone score; among others. These clinical parameters will be added and tested within the predictive model in addition to radiomics markers to evaluate their contributions for predicting bone fractures.

2.3. Data Processing

2.3.1. Segmentation and mask generation

Fig. 3. a) original data provided by CETIR, automatic segmentation approach contouring the femur and pelvis bone. b) Correction of the original segmentation approach delineating femur and pelvis bone. c) Segmentation of the femur bone.

Segmentation was carried out to define the femur boundaries within which the radiomics features will be calculated. A first automatic segmentation approach was provided by CETIR following the femur and pelvis boundaries, but these segmentations were not entirely correct at some points of the image so manually corrections was performed (see
Fig. 3. a and b). Once the scans were properly segmented, masks were created from the segmentations as to indicate the femur and pelvis regions (see Figs. 3 and 4). Note that both the femur alone, as well as the femur together with the pelvis, will be tested in this study to identify the most suitable bone structures for predicting risk fracture (i.e. femur alone or femur + pelvis; see section 3.2: Classification results per bone structure). The segmentation of the femur bone was obtained by manually delineating the head of the femur (see Fig. 3. c) to remove the pelvis bone, as shown by the masks obtained in Fig. 4.

![Fig. 4. a) DXA of the segmented femur visualization with ITK Snap tool. b) DXA of the segmented femur with its corresponding mask visualization using ITK Snap tool.](image)

### 2.3.2. Registration process

The purpose of registration was to map specific regions within the femur (a total of regions instead of using the entire femur, see Fig. 5). This more localized approach is motivated by the existing studies which showed that there exist specific high-risk regions of the femur where fractures take place [29], [39]. For example, López et al. [29] computed fracture probability maps to identify those zones more susceptible to suffer a fracture, which resulted in regions corresponding to the femoral neck, intertrochanteric and subtrochanteric areas.

In this study, only these specific zones of interest inside the femur were chosen to extract radiomics data. Dr. Carlos Ruiz Wills (Multiscale and Computational Biomechanics and Mechanobiology research team, Universitat Pompeu Fabra) conducted a biomechanical study (to be published) based on the subdivision of the femur described above and illustrated in Fig. 5. The same subdivision is used in this study and propagated to our dataset by using image registration. Registration is a common used technique in medical imaging analysis, which consists on transforming an image into another the coordinate
system of another image. In the present study, the original images from CETIR were defined as fixed images (reference coordinate system) and the image containing the femur subdivisions of interest from the 3D model was set as the moving image (the one being transformed to fit into the reference image) (Fig. 5). With this approach, the subdivisions are propagated to all the 65 cases of our dataset (see Fig. 6).

![Fig. 5. a) DXA segmentation of the femur, corresponding to the fixed image in the registration process. b) Image containing regions of major interest, corresponding to the moving image in the registration process.](image)

The registration process itself consisted of first binarizing both fixed and moving images (i.e. inside and outside of the femur), considering that intensity values were not important for this case. Subsequently, an affine transformation is used for the registration, including translation, rotation, scaling and skewing. The registration process was performed using the `imregtform` function defined in MATLAB. Note that despite of the fact that mostly the neck and the union of trochanters regions were defined as of major interest previously, it was considered that extracting indices of all subdivisions could add value to the algorithm as additional features would be added to the model. In other words, five zones were defined, registered to all DXA images and their corresponding masks were created to perform radiomics over them (Fig. 6.b). The definition of the five regions were the following ones:

- Femoral neck (orange)
- Neck of the femur (blue)
- Intertrochanteric line (green)
- Union of greater and lesser trochanter (red)
- Pectineal line (pink)
In this work, the proposed predictive modeling of bone fracture will be carried out based on the obtained region-based subdivision of the femur as explained above. Thus, the radiomics features will be calculated in each of these regions separately and analyzed by considering the following regions alone or combined (see Section 3):

1. Neck of the femur (blue region in Fig. 6a) combined with the area connecting the greater with the lesser trochanter (red region in Fig. 6a).
2. Neck of the femur (blue) combined with the area representing the union of the trochanters (red) and the intertrochanteric region (green).
3. Combination of all regions.

Furthermore, two different registration accuracy metrics are calculated and tested in this work, namely the Dice Coefficient (DC, which measures the degree of overlap between the obtained masks and those provided as ground truth) and the Correlation Coefficient (CC, used to measure the strength of the relationship between the intensity distributions of the fixed and moving bone images). Both metrics range between 0 (samples are totally different) and 1 (samples are in full agreement). The results are provided in Section 3.1.

### 2.4. Input of the predictive model of fracture risk

Once the segmentations and registrations are performed, resulting in the definition of all the regions of interest within the bones for all cases, the next step is to estimate the input
of the predictive model, namely the radiomics features (Section 2.4.1) and the clinical variables (Section 2.4.2).

2.4.1. Radiomics features for fracture prediction

Radiomics feature extraction was implemented in Python 3.6 by using the open-source package PyRadiomics [40]. The obtained features describe properties of the bone tissues they describe, including size, shape, intensity and texture-based, as detailed in this section.

Shape features: They include measures such as area, diameters, orientation, elongation, circularity, etc.

First order intensity statistics: Include measures describing the distribution of pixel intensities within the regions of interest defined by the masks. These include mean intensity, range, skewness (i.e. asymmetry of distribution), etc.

Grey Level Co-occurrence Matrix (GLCM) features: describe the second-order joint probability function of the established zone of interest, in other words the repeatability of intensity values. The values of the obtained matrix \((i,j)\) represent the number of times the combination of levels \(i\) and \(j\) occurred in a pixel separated from a certain distance and given a specific angle.

Grey Level Size Zone Matrix (GLSZM) features: This class of radiomics quantifies zones of grey level from an image, where grey level zones correspond to connected pixels with the same intensities.

Grey Level Run Length Matrix (GLRLM) features: They quantify consecutive pixels with the same grey-level intensity, where the level run corresponds to the number of pixels in length with the mentioned characteristics. The obtained matrix represents the number of runs with grey level \(i\) and length \(j\) occurring within the zone of interest.

Neighboring Grey Tone Difference Matrix (NGTDM) features: This class of radiomics indices quantify the difference between a grey value and the average grey value of its neighbors given a given distance. This it describes homogeneity/heterogeneity of the tissue appearance.

Grey Level Dependence Matrix (GLDM) features: These features are useful for quantifying grey-level dependencies or number of connected voxels in a certain distance from an image that are dependent on the center voxel.
2.4.2. Clinical variables

In this study, in addition to the radiomics calculated in previous section as the input for our predictive model of fracture, clinical characteristics of the patients will be added with the goal to obtain an even more accurate prediction. For this purpose, the selected data to add to the predictive model include age, gender, weight, height, Body Mass Index, body fat, hip tissue thickness, Mean Trabecular Bone Score, previous fragility fracture and family history fracture. All these parameters have been analyzed in many studies, in which they were found to be risk factors for osteoporotic related fractures [26]. Finally, diabetes was also added as it was reported even though it is not a major risk factor. Table 1 summarizes the clinical variables used in this thesis and their numerical properties.

Table 1. List of clinical variables tested within the classification model.

| Names and characteristics of the clinical variables within the study |  
|---|---|
| Age (mean ± S.D) | 79.5 ± 7.8 (years) |
| Weight (mean ± S.D) | 70.6 ± 13.1 (kg) |
| Height (mean ± S.D) | 1.55 ± 7.3 (m) |
| Body Mass index (mean ± S.D) | 29.5 ± 5.9 (kg / m²) |
| Body Fat (mean ± S.D) | 38.9 ± 5.7 |
| Hip tissue thickness (mean ± S.D) | 18.6 ± 2.7 |
| Mean Spine Trabecular bone score (mean ± S.D) | 1.2 ± 0.1 |
| Gender n (%) |  
| Women | 46 (74%) |
| Men | 16 (26%) |
| Site of fracture (over the 30 patients with fracture) n(%) |  
| Femoral neck | 13 (43.33%) |
| Trochanter | 16 (53.33%) |
| Unknown | 1 (3.33%) |
| Previous fragility fracture n (%) | 2 (3%) |

2.5. Construction of the predictive model

Two final steps are required to build the predictive model of femur risk, namely normalization of the input data (Section 2.5.1) and implementation of suitable machine learning methods for classification (Section 2.5.2).
2.5.1. Normalization of the input of the predictive model

The radiomics which will be used as the input of the predictive model are of different characteristics, units and scales. Therefore, they need to be normalized before they are entered into the predictive model. Similarly, the clinical variables (e.g. age, gender, weight), which have different units and are also different to the radiomics units, will be normalized as well. The goal of this normalization is to ensure all input variables contribute equally to the model and thus to eliminate potential bias.

This was achieved by using the `scikit.learn preprocessing` package available for Python. The tested normalized methods include:

**Standard Scaler:** For each radiomic or clinical variables, this method standardizes the values by removing the mean and scaling to unit variance.

**Robust Scaler:** This method is more robust to the presence of outliers. It scales the features by removing the median and scaling to the interquartile range (from the 25th quantile to the 75th).

**MinMax Scaler:** This pre-processing method transforms the between zero and one by translating and scaling each feature based on the min and max values.

**UnitNorm Scaler:** This method is used to scale all individual samples as to have unit norm. It is often used for clustering and classification problems in which the similarity between two samples is quantified by kernels (see Section 2.5.2).

Otherwise stated, the default normalization for the experiments in Section 3 were obtained with the RobustScaler method.

2.5.2. Machine learning models

Three machine learning algorithms were tested in this study to build the predictive models [41].

**Support Vector Machines (SVM):** This algorithm consists of finding a hyperplane that best separates the different classes in the feature space. For a new case under investigation, it is mapped into the feature space and classified based on its position with respect to the hyperplane. As there might be many possible solutions, the selected hyperplane is the one having the maximum distance to the closer sample from each class, the so-called maximum-margin hyperplane. In case the classes are not linearly separable
in the given space, a kernel function of the input data is performed to remove the non-linearity. Note that a kernel is a function that projects each of the original data points into a new vector space (higher dimensional) [42]. In this study, three different kernels are compared:

1. **Linear kernel function**: It is the fastest of all kernel transformations. It is useful and accurate when the number of features is large.
2. **Radial basis kernel function, or Gaussian kernel**: It can be interpreted as a similarity measure, which transforms the dot product in the original feature space (nonlinear) into a Gaussian dot product in the linearized space.
3. **Sigmoid kernel function**: Its origin comes from the field of neural networks and is often tested as an alternative to the Gaussian kernel [43].

**Decision Trees.** This method classifies categorical data hierarchically based on the value of their attributes and by using predefined thresholds at each branch of the classification tree. The algorithm takes decisions at different nodes, such that at each of this stage an attribute is selected to split the dataset into the different classes. From these nodes, branches are created ending in other nodes where data is further classified based on another attribute. The procedure consists on firstly selecting the best attribute at the “root” of the tree (whole training set), then splitting it into subsets (new branches) containing data with same-class values (preferably categorical values, if not discretization is needed) based on the selected attribute. The procedure is repeated at subsequent branches, which will split into branches in a similar fashion, until finding the so-called leaf nodes for each branch (the end branches) [44].

**Random Forest**: Also called Random Decision Forest, this method consists of using multiple learning models to improve predictability. Concretely, this algorithm builds many decision trees with the training set and gives a prediction based on the mode of the labeled classes obtained from all generated trees. Trees are grown as independent bootstrap samples using regeneration of the samples at random. All trees are extended as deeper as possible; this means there is no pruning. Random Forest is fast and does not tend to overfit the model [45].
2.5.3. **Feature Selection**

In Section 2.3.1, a total of 525 radiomic features are calculated for the needs of this study. However, including all these variables to the predictive model could lead to the so-called overfitting problem given the limited sample size used in this study (63 cases). To resolve this problem, we propose to perform feature selection, which consists of selecting the most relevant variables. The process of selecting features eliminates non-informative and redundant features, ensuring robust and reproducible analyses. Two different feature selection techniques pertaining to the family of the greedy search methods have been used within the project:

**Sequential Forward Feature Selection (SFFS):** With this approach, the optimal feature set is initiated with an empty set. Subsequently, and iteratively, the next feature associated with an improvement of the classification accuracy in cross-validation tests is added to the set of optimal features, until no improvement in the classification is noticed.

**Sequential Backward Feature Selection (SBFS):** In this method, the set of optimal features is initiated with the whole set of features. Subsequently, and iteratively, features are removed from the set one at a time such that at each removal the classification improves best. SBFS can be seen as the inverse process of the SFFS.

Otherwise stated, we select in the experiments 10 features.

2.6. **Model validation**

The final part of this thesis is to estimate the accuracy of the predictive models using cross-validation tests. To do so, the whole dataset is split into 10 subsets, each one playing the role of the test data sequentially, while the rest of the cases are used for training.

The metric used to test the algorithm is the classification accuracy, which is equals the number of correct classifications divided by the total number of cases.
3. RESULTS AND DISCUSSION

3.1. Evaluation of the registration process
As a first step of the validation, the performance of the registration process was evaluated. This is important to assess that the definitions of the subdivisions within each femur is suitably performed before the radiomics are calculated. The obtained results are as follows:

Dice coefficient calculated over registration results: 0.928
Cross-correlation calculated over registration results: 0.915

These results, close to 1, indicate strong agreement between the automatically obtained results and the ground truth.

3.2. Classification results per bone structure
In this section, we compare the classification results achieved by the proposed predictive models for different bone structures, namely the femur and pelvis bones. The goal is to test whether fracture can be predicted more accurately by focusing on specific bones. As it can be observed in Table 2, integrating the pelvis into the predictive model reduces the classification accuracy, while the predictive model with the femur alone provides the highest accuracy. As it can be seen from Figs. 4-6, the geometry and appearance of the pelvis bone is more complex which leads to reduced accuracy.

Table 2. Classification accuracy for features extracted from different bone structures (femur and pelvis bones).

<table>
<thead>
<tr>
<th></th>
<th>SVM (linear kernel)</th>
<th>SVM (rbf kernel)</th>
<th>SVM (sigmoid kernel)</th>
<th>Random Forest</th>
<th>Decision Tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Femur and pelvis</td>
<td>0.7619</td>
<td>0.7333</td>
<td>0.7476</td>
<td>0.7738</td>
<td>0.7310</td>
</tr>
<tr>
<td>Femur alone</td>
<td>0.7381</td>
<td>0.8143</td>
<td>0.7833</td>
<td>0.7929</td>
<td>0.8429</td>
</tr>
</tbody>
</table>

Based on this result, the rest of the evaluations in subsequent sections is performed by focusing on the femur. The goal now is to optimize the parameters of the algorithm in order to improve over the 0.8429 classification accuracy.
3.3. Results per region within the femur

In this section, the predictive model was built and evaluated for different sub-regions of the femur.

Table 3. Classification accuracy by using feature combinations extracted from different regions

<table>
<thead>
<tr>
<th>Region</th>
<th>SVM (linear kernel)</th>
<th>SVM (rbf kernel)</th>
<th>SVM (sigmoid kernel)</th>
<th>Random Forest</th>
<th>Decision Tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neck and trochanters’ union regions</td>
<td>0.7333</td>
<td>0.8595</td>
<td>0.7905</td>
<td>0.7786</td>
<td>0.7905</td>
</tr>
<tr>
<td>Neck, trochanters’ union and intertrochanteric regions</td>
<td>0.7619</td>
<td>0.8595</td>
<td>0.7905</td>
<td>0.7476</td>
<td>0.7929</td>
</tr>
<tr>
<td>All regions</td>
<td>0.8571</td>
<td>0.8881</td>
<td>0.8095</td>
<td>0.7357</td>
<td>0.8714</td>
</tr>
</tbody>
</table>

Table 3 shows that by estimating radiomics from the regions of interest within the femur, the accuracy improved compared to features extracted from the whole femur. When using biomarkers from the whole femur the best performance was 0.8429, while when combining regions defined of high-risk of hip fracture (neck and trochanters’ union) was 0.8595. However, it was observed that adding the indices from the intertrochanteric region gave no meaningful information to the algorithm. Finally, when combining features from all regions the accuracy was surpassed, obtaining 0.8881.

Using specific regions clearly improved the fracture probability outcome with respect of using the entire bone. Moreover, it has been proven that concretely the femoral neck and the union of greater and lesser trochanters regions have predictability potential for fractures at the hip.

Despite of the acceptable accuracy reached when using the two zones, the algorithm improved its result when markers from all regions were combined. This fact suggests that the quantification of hidden information within the bone tissue gives better accuracies when extracted locally.
3.4. Impact of the number of radiomic features

Up to this point, a promising accuracy of 0.8881 was obtained by using in each test a combination of 10 features as input for the predictive models. In this section, we test the impact of varying the number of features on the classification, from a single radiomic feature to the maximum of 525 features. The forward sequential forward selection method was used to guide the feature selection and the machine learning technique used in this experiment was SVM. The classification accuracy is plotted in Fig. 7 as a function of the number of radiomics features selected and used in the predictive model.

![Graph](image)

**Fig. 7.** Accuracy of the predictive model as a function of the number of features selected in the model.

It can be seen that the maximum accuracy was reached by selecting 13 radiomics features, namely 0.9023. Afterwards, the addition of new radiomics features not only does not improve the fracture prediction, but in facts reduces it as the model becomes overfitted due to the limited sample size, falling below 0.80 when using 50 features.
3.5. Incorporating clinical variables to the predictions

Table 4. Classification accuracy including clinical variables to the input of the model

<table>
<thead>
<tr>
<th></th>
<th>SVM (linear kernel)</th>
<th>SVM (rbf kernel)</th>
<th>SVM (sigmoid kernel)</th>
<th>Random Forest</th>
<th>Decision Tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>With clinical variables</td>
<td>0.7405</td>
<td>0.8190</td>
<td>0.7929</td>
<td>0.7357</td>
<td>0.8238</td>
</tr>
<tr>
<td>Without clinical variables</td>
<td>0.8571</td>
<td>0.8881</td>
<td>0.8095</td>
<td>0.7357</td>
<td>0.8714</td>
</tr>
</tbody>
</table>

In this section, an experiment is performed to test the impact of adding clinical variables to the predictive model. Thus, the feature selection was performed not only on the 525 radiomics features but on a total of 536 indices integrating the 11 clinical risk factors listed in Table 1. However, it can be seen from the results given in Table 4 that the predictive model loses accuracy in this case (classification accuracy = 0.8238), while by focusing on the imaging variables (i.e. the radiomics) the performance is improved to 0.8881. This result suggests that the machine learning model has a better behavior when considering imaging features only and that its behavior is affected by adding heterogeneous data types. This behavior is particularly common in the field of machine learning when working with relatively small training samples, as there is limited evidence to model the more complex heterogeneous variability.

3.6. Analysis of radiomics types

Table 5. Classification accuracy of features from different classes

<table>
<thead>
<tr>
<th></th>
<th>SVM (linear kernel)</th>
<th>SVM (rbf kernel)</th>
<th>SVM (sigmoid kernel)</th>
<th>Random Forest</th>
<th>Decision Tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape features</td>
<td>0.7548</td>
<td>0.7024</td>
<td>0.6905</td>
<td>0.6929</td>
<td>0.6952</td>
</tr>
<tr>
<td>Intensity features</td>
<td>0.6810</td>
<td>0.7333</td>
<td>0.7310</td>
<td>0.6643</td>
<td>0.8095</td>
</tr>
<tr>
<td>Texture features</td>
<td>0.8262</td>
<td>0.7476</td>
<td>0.7810</td>
<td>0.7071</td>
<td>0.7762</td>
</tr>
<tr>
<td>All features</td>
<td>0.8571</td>
<td>0.8881</td>
<td>0.8095</td>
<td>0.7357</td>
<td>0.8714</td>
</tr>
</tbody>
</table>
As explained in the Method Section, three classes of radiomics features are used as input for the classification model in this work, namely shape, intensity and texture based. In this subsection, the impact of each class of features is analyzed by performing the feature selection for each of the radiomics classes separately, selecting the 10 best feature in each test.

From the results in Table 5 it is concluded that quantitative texture indices are the best class of radiomics indices for the task of predicting fracture (accuracy 0.82, against 0.75 for the shape radiomics and 0.80 for the intensity radiomics). This is an expected result as the texture of the bone tissue provides richer information of the local state of the bone tissue and is thus more indicative of fracture risk than shape-based radiomics for example. However, the results also show the importance of combining features of different nature, as they are complimentary to each other and enrich the evidence given to the prediction model. By adding shape and intensity based features in addition to texture radiomics, the accuracy improves 0.82 to 0.88.

3.7. Impact of the feature selector

Up to this point, the best accuracy was obtained by selecting 13 features using sequential forward feature selection in Section 3.4. In this section, the Sequential Backward Feature Selection (SBFS) method was used as the feature selector instead of the Forward Selection approach (SFFS). Note that the SFFS was the method tested in all section because it is less time consuming than SBFS. However, its main drawback is that it is unable to remove features that become obsolete once they have been selected. Furthermore, SBFS works best when using a large number of features which is the case of this study.

In this case, it can be seen from the results in Figure 8 that the performance of the SBFS algorithm improves compared to the one achieved for SFFS. Using SVM with a linear kernel as the learning model and SBFS, an accuracy equal to 0.9857 was achieved by selecting 26 radiomics features. Furthermore, a maximal accuracy of 1.0 was reached when selecting 30 features, meaning the predictive model correctly classified all cases.
3.8. Best Selected Features

Finally, a detailed analysis of the radiomic indices selected as the optimal ones for the fracture prediction is performed in this section. Table 6 lists the optimal radiomics, together with their class and the femur sub-region they belong to. Furthermore, for each feature, the accuracy values of the predictions by removing it from the model or by using it alone are also provided to understand the importance of each radiomic feature. Firstly, it can be observed that the optimal selected features listed include 17% of shape, 20% of intensity and 63% texture features. These results coincided with the ones obtained in the analysis of radiomics features classes in Section 3.6, where the texture radiomics provided the highest accuracy.

<table>
<thead>
<tr>
<th>Feature Name</th>
<th>Type</th>
<th>Region</th>
<th>Without</th>
<th>Alone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum 2D Diameter Column</td>
<td>Shape</td>
<td>Neck</td>
<td>0.8095</td>
<td>0.5381</td>
</tr>
<tr>
<td>Maximum 2D Diameter Row</td>
<td>Shape</td>
<td>Neck</td>
<td>0.7976</td>
<td>0.5738</td>
</tr>
<tr>
<td>Maximum 2D Diameter Slice</td>
<td>Shape</td>
<td>Neck</td>
<td>0.9071</td>
<td>0.5071</td>
</tr>
<tr>
<td>Median</td>
<td>Intensity</td>
<td>Neck</td>
<td>0.9548</td>
<td>0.5857</td>
</tr>
<tr>
<td>Robust Mean Absolute Deviation Correlation</td>
<td>Intensity</td>
<td>Neck</td>
<td>0.9690</td>
<td>0.5095</td>
</tr>
<tr>
<td>Low Gray Level Zone Emphasis</td>
<td>Texture</td>
<td>Neck</td>
<td>0.9714</td>
<td>0.5214</td>
</tr>
<tr>
<td>Size Zone Non-Uniformity Normalized</td>
<td>Texture</td>
<td>Neck</td>
<td>0.9381</td>
<td>0.4905</td>
</tr>
<tr>
<td>Small Area High Gray Level Emphasis</td>
<td>Texture</td>
<td>Neck</td>
<td>0.8571</td>
<td>0.5738</td>
</tr>
<tr>
<td>Strength</td>
<td>Texture</td>
<td>Neck</td>
<td>0.8262</td>
<td>0.5214</td>
</tr>
<tr>
<td>LeastAxis</td>
<td>Shape</td>
<td>Trochanters’ union</td>
<td>0.8881</td>
<td>0.4881</td>
</tr>
<tr>
<td>Feature Class</td>
<td>Region</td>
<td>10Percentile</td>
<td>Intensity</td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>---------------------</td>
<td>--------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>10Percentile</td>
<td>Intensity</td>
<td>0.9190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kurtosis</td>
<td>Intensity</td>
<td>0.8214</td>
<td>0.5571</td>
<td></td>
</tr>
<tr>
<td>Cluster Prominence</td>
<td>Texture</td>
<td>0.9500</td>
<td>0.5071</td>
<td></td>
</tr>
<tr>
<td>Contrast</td>
<td>Texture</td>
<td>0.8429</td>
<td>0.5214</td>
<td></td>
</tr>
<tr>
<td>Low Gray Level Zone Emphasis</td>
<td>Texture</td>
<td>0.8262</td>
<td>0.5976</td>
<td></td>
</tr>
<tr>
<td>Size Zone Non-Uniformity Normalized</td>
<td>Texture</td>
<td>0.9048</td>
<td>0.5048</td>
<td></td>
</tr>
<tr>
<td>Least Axis</td>
<td>Shape</td>
<td>0.9048</td>
<td>0.4929</td>
<td></td>
</tr>
<tr>
<td>Imc2</td>
<td>Intertrochanteric</td>
<td>0.8738</td>
<td>0.4905</td>
<td></td>
</tr>
<tr>
<td>Large Dependence Low Gray Level Emphasis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10Percentile</td>
<td>Intensity</td>
<td>0.9690</td>
<td>0.5738</td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>Intensity</td>
<td>0.7952</td>
<td>0.5380</td>
<td></td>
</tr>
<tr>
<td>Contrast</td>
<td>Texture</td>
<td>0.9690</td>
<td>0.4286</td>
<td></td>
</tr>
<tr>
<td>Idn</td>
<td>Texture</td>
<td>0.8381</td>
<td>0.5833</td>
<td></td>
</tr>
<tr>
<td>High Gray Level Zone Emphasis</td>
<td>Texture</td>
<td>0.9357</td>
<td>0.5857</td>
<td></td>
</tr>
<tr>
<td>Large Area High Gray Level Emphasis</td>
<td>Texture</td>
<td>0.9571</td>
<td>0.5238</td>
<td></td>
</tr>
<tr>
<td>Size Zone Non-Uniformity Normalized</td>
<td>Texture</td>
<td>0.8429</td>
<td>0.4119</td>
<td></td>
</tr>
<tr>
<td>Difference Variance</td>
<td>Texture</td>
<td>0.8762</td>
<td>0.5357</td>
<td></td>
</tr>
<tr>
<td>Gray Level Variance</td>
<td>Texture</td>
<td>0.9024</td>
<td>0.4714</td>
<td></td>
</tr>
<tr>
<td>Size Zone Non-Uniformity Normalized</td>
<td>Texture</td>
<td>0.8762</td>
<td>0.5381</td>
<td></td>
</tr>
</tbody>
</table>

Furthermore, the selected features are distributed among all femur regions, indicating that evidence from the entire femur is needed to identify potentially fracture-prone areas. However, the regions providing the most relevant information to the model are from the femoral neck zone with 33% of features, followed by trochanter’s union and the regions with 23% of features each. These findings are in agreement with the existing knowledge that the neck and the union of trochanters regions are of higher relevance as the areas at which femur fracture occurs most.

The exclusion of specific features from the model had a negative impact over the obtained accuracy, which decreased from 1.0 to 0.7952 by eliminating the so-called Minimum feature from the femoral head region, though its own this feature only achieves a 0.53 accuracy.

It can also be seen that on separately, radiomics indices are not able to perform a satisfactory classification, with the single-feature accuracies ranging between 0.5976 as the best result obtained using the Low Grey Level Emphasis from the trochanters’ union region to 0.4119 from the Size Zone Non-Uniformity Normalized feature from the head of the femur zone. This last feature mentioned, with the lowest accuracy of 0.4119 cannot
produce any meaningful classification alone. However, its contribution to the model is important to the overall accuracy, as when removing it from the overall model the accuracy is reduced from 1.0 to 0.84. These results further show the complementarity of the different radiomics features and types and how their combination is critical to achieve optimal predictions of bone fracture.

4. CONCLUSIONS

Summary of findings. From the different experiments achieved in the Validation Section, the following conclusions can be made:

1. Fracture prediction is best achieved by integrating radiomics from the femur only, as the pelvis includes complex appearance that is detrimental to the model.
2. All sub-regions of the femur are important for achieving accurate prediction of bone fracture, as they provide complimentary evidence about bone’s state. However, the neck and the union of trochanters regions are those that lead to the best predictions in agreement with the existing clinical knowledge.
3. Similarly, all radiomics types contribute to fracture prediction and they provide complimentary evidence to the model. However, texture-based radiomics are the most accurate as they inform the model about the bone’s local state.
4. The best approach for selecting the optimal radiomic features is the SBFS. While it is slower than SFFS, it can be run offline at the training stage, thus it does not affect the application of the predictive model in clinical practice.
5. The inclusion of clinical variables in addition to the radiomic features makes the predictive model more heterogeneous and its accuracy is reduced. Given the limited sample size, a predictive model focused on imaging evidence through radiomics results in a more stable behavior, as well as fracture prediction.
6. Finally, by using SVM as the machine learning model, combined with a total of 30 radiomics selected through SBFS, we obtain a perfect classification accuracy of 1.0.

Limitations and further work. This work has some limitations that are important to mention. First, while the sample size consisting of 63 cases enabled us to perform a wide range of experiments and test different situations, with subjects providing sufficient variability (age, gender, weight, etc), it remains relatively limited to draw definitive conclusions on the predictive capacity of the proposed radiomics-based model.
Additional multi-center validations are required to test applicability in diverse clinical populations. Furthermore, the segmentation process in this thesis proved challenging, and was achieved with the aid of an expert observer to ensure the regions of interest are properly defined for the extraction of the radiomics markers. The development of more advanced and accurate segmentation techniques such as using the deep learning paradigm would further enhance the construction and application of the predictive model of fracture risk.

Finally, it is important to note that the regions of interest (ROIs) of the femur defined in this study do not completely coincide with those used in clinical practice, especially the distinction between the head and the femoral neck. Currently, each DXA manufacturer has its own lexicon to identify the ROIs, which are not completely similar. As such, one important future direction is to find a more suitable subdivision of the femur such that are be located in the same anatomical coordinates across samples and must offer greater performance (greater sensitivity / specificity) than the regions used so far. For example, Dr. Luis Miguel del Río from CETIR Medical Center proposed new ROIs taking into account the location of the maximum stresses in falls simulations (see Fig. 9) with the aim of finding a better predictive and diagnostic performance.

In summary, this thesis represents the very first attempt to build a rich predictive model of bone fracture from DXA by using radiomics to encode subtle as well as advanced characteristics of the bone tissues. These preliminary results show the promise of the approach and promote future works in this field. Note that a journal paper summarizing these results is in preparation and will be submitted in the next few months to the journal BONE.

**Fig. 9.** New regions of interest for radiomics-based prediction of bone fracture.
5. BIBLIOGRAPHY


