الجمهورية الجزائرية الديمقراطية الشعبية
Republique Algerienne Democratique Et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة العربي التبسي – تبسة
Université Larbi Tébessi – Tébessa –
Faculté des Sciences et de la Technologie
Département de Génie Electrique

MEMOIRE

Présenté pour l'obtention du **diplôme** de **Master Académique**

En : Automatique

Spécialité : Automatique et systèmes

Par: DJEDOUANI Nedjmeddine and HAFSI Haithem

Sujet

Détection automatique de l'ostéoporose à travers des images radiographiques utilisant des caractéristiques profondes

Evaluée le 22/06/2021 par :

M. AMIEUR Toufik	MCA	Président
M. MERAOUMIA Abdallah	MCA	Rapporteur
M. DJABRI Riad	MCB	Examinateur

Promotion : 2020/2021

الجمهورية الجزائرية الديمقراطية الشعبية

People's Democratic Republic of Algeria

وزارة التعليم العالي والبحث العلمي

Ministry of Higher Education and Scientific Research

جامعة العربي التبسي – تبسة

Larbi Tebessi University – Tebessa –

Faculty of des Science and Technology

Electrical Engineering Department

DISSERTATION

Presented for obtaining the **Academic Master degree**

In: Automatic

Specialty: Automatic and systems

By: DJEDOUANI Nedjmeddine and HAFSI Haithem

Theme

Automatic detection of osteoporosis through X-ray images using deep features

Evaluated the 22/06/2021 by :

Mr. AMIEUR Toufik	MCA	President
Mr. MERAOUMIA Abdallah	MCA	Rapporteur
Mr. DJABRI Riad	MCB	Examinator

Academic year: 2020/2021

Résumé *: L'ostéoporose est une maladie osseuse prédominante, qui fragilise l'os et est l'un des principaux facteurs de handicap, en particulier chez les personnes âgées. Elle se caractérise généralement par une masse osseuse à faible densité minérale et une détérioration micro-architecturale du tissu osseux. Dans ce mémoire, nous proposons un système efficace de détection de l'ostéoporose utilisant les caractéristiques profondes des images. Dans cette étude, une image d'ostéoporose segmentée est d'abord divisée en blocs superposés et de taille égale, puis applique la méthode d'extraction de caractéristiques PCANet à chaque bloc. En utilisant le processus de concaténation, toutes les caractéristiques des blocs sont combinées pour produire le vecteur de caractéristiques. Par la suite, nous utilisons la technique de sélection de caractéristiques pour sélectionner les caractéristiques dominantes. Enfin, nous utilisons le classifier SVM pour classer le vecteur de caractéristiques de chaque image de test. La méthode proposée est validée pour son efficacité sur une base de données typique et disponible de l'ostéoporose et nos résultats expérimentaux montrent l'efficacité et la fiabilité de la méthode proposée pour détecter la maladie et ensuite améliorer la capacité des experts à déterminer le risque d'ostéoporose lors de l'examen des patients.*

Mots clés : *Ostéoporose, Extraction de caractéristiques, Apprentissage profond convolutif, Sélection de caractéristiques, PCANet*

Abstract: *Osteoporosis is a prevailing bone disease, which weakens the bone and is one of the major factors of disability, especially in elderly persons. It is generally characterized by low bone mineral density mass and micro-architectural deterioration of bone tissue. In this dissertation, we propose an efficient osteoporosis detection system using the deep features of images. In this study, a segmented osteoporosis image is first divided into overlapping and equal-sized blocks and then applies the PCANet feature extraction method to each block. Using the concatenation process, all the blocks' features are combined to produce the feature vector. Subsequently, we use the feature selection technique to select the dominant features. Finally, we use SVM to classify the feature vector of each osteoporosis test image. The proposed method is validated for its effectiveness on a typical and available database of osteoporosis and our experimental results show the effectiveness and reliability of the proposed method to detect the disease and then improve the ability of the experts to determine the risk of osteoporosis during patient's examination.,*

Index term*: Osteoporosis, Feature extraction, Convolutional deep learning, Feature selection, PCANet.*

ملخص: هشاشة العظام من أمراض العظام السائدة التي تضعف العظام وتعتبر من العوامل الرئيسية للإعاقة خاصة عند كبار السن. يتميز بشكل عام بانخفاض كثافة المعادن في العظام وتدهور معماري دقيق لأنسجة العظام. في هذا البحث، نقترح نظامًا فعالاً للكشف عن هشاشة العظام باستخدام الميزات العميقة للصور. في هذه الدراسة، يتم أولاً تقسيم صورة العظام المجزأة إلى كتل متداخلة ومتساوية الحجم ثم يتم تطبيق طريقة استخراج ميزة PCANetعلى كل كتلة. باستخدام عملية التسلسل، يتم دمج جميع ميزات الكتل لإنتاج متجه الميزة. بعد ذلك، نستخدم تقنية اختيار الميزة لتحديد الميزات السائدة. أخيرًا ، نستخدم SVM لتصنيف متجه الميزات لكل صورة اختبار هشاشة العظام. تم التحقق من صحة الطريقة المقترحة لفعاليتها على قاعدة بيانات نموذجية ومتاحة لهشاشة العظام وتظهر نتائجنا التجريبية فعالية وموثوقية الطريقة المقترحة للكشف عن المرض ومن ثم تحسين قدرة الخبراء على تحديد مخاطر الإصابة بهشاشة العظام أثناء فحص المرضى. .

الكلمات المفتاحية هشاشة العظام ، استخراج الميزات ، التعلم العميق التلافيفي ، اختيار الميزات ، PCANet.

Acknowledgment

After the praise to ALLAH the Almighty, who gave us the strength and patience to complete this Modest work, we would like to express our sincere thanks and gratitude to our supervisor Dr. Abdallah Meraoumia, for the continuous support in this work, for his motivation, and immense knowledge. His guidance helped us in all the time of writing of this thesis.

We would like to thank Prof. Seddik Khemaissia and PhD student miss Meriem Mebarkia for their assistance and willingness to provide feedback, which made the completion of this work an enjoyable experience.

And to everyone who helped us in this work, each in his name, to you all the sincerest expressions of thanks.

H. Haithem and J. Nedjmeddine

Dedication

I have the great pleasure to dedicate this work to:
My parents
For all the encouragement they gave me despite their heavy responsibility
May god inchaa Allah keep them for us
My brothers and sisters
My supervisor **Abdalllah Meraoumia**
who have supported us throughout this work
All teachers of the electrical engineering department
All the Class of Control Systems
My friends and particularly the most intimate ones
Yacine and **Rafik** for the unforgettable moments, and the strong bonds that unite us
Please accept my best wishes for success and prosperity

Djedouani Nedjmeddine

This dissertation is dedicated to:
Those who are impossible to thank them adequately, for everything they have done, for their endless support and encouragement throughout my life,
My Dear Parents
I could not have asked for better parents or role models.
And to my precious brother and sister and to my little sweet niece and nephew
To my supervisor **Abdallah meraoumia**
Also to all my classmates and teachers in my study career
And to my friends

Hafsi Haithem

Table of Contents

Glossary

The following abbreviations, listed in alphabetical order, are used in the text.

ACC : Accuracy.
AUC : Area under curve.
BMD : Bone Mineral Density.
BoVW : Bag of visual words.
CAD : Computer Aided Design system.
CT : Computed tomography.
DXA : Dual Energy X-ray Absorptiometry.
FAST : Features from Accelerated Segment Test.
FDA : Food and Drug Administration.
FPR : False positive rate.
GLCM : Gray Level Co-occurrence Matrix.
IEEE : Institute of Electrical and Electronics Engineers.
ISBI : International Symposium on Biomedical Imaging.
LBP_{riu2} : Improved Local Binary Pattern.
MRI : Magnetic Resonance Imaging.
ORB : Oriented FAST and Rotated BRIEF.
RBF : Radial Basis Function.
ROC : Receiver Operating Curve.
ROI : Region of interest.
SIFT : Scale Invariant Feature Transformation.
SVM : Support vector machine.
TPR : True positive rate.

List of Figures

List of Tables

General Introduction

Introduction

OSTEOPOROSIS is the most common bone disease in humans, representing a major public health problem. This disease affects the skeleton and is characterized by low bone mass density and micro-architecture deterioration of the bone tissue, with a consequent increase in bone fragility and susceptibility to fracture. Thus, early detection and treatment of osteoporosis can decrease the fracture risk of a person to the minimum. A lot of studies have been undertaken to identify clinical risk factors that could be used. The diagnosis of osteoporosis is based primarily on the measurement of bone mineral density (BMD). BMD tests can identify osteoporosis by evaluating bone mineral density (BMD) measurements (expressed as a T-score) using dual-energy X-ray absorptiometry (DXA), which is considered as the reference standard examination for BMD assessment. However, BMD alone represents only 60% of fracture prediction. The characterization of the trabecular bone micro-architecture has been recognized as an important factor and completes the diagnosis of osteoporosis using BMD. The micro-architecture of the bone cannot be routinely obtained by noninvasive methods and requires a bone biopsy with histomorphometric analysis. This is why several attempts have been made to characterize the trabecular bone micro-architecture by noninvasive methods.

An image-based medical diagnostic system is typical pattern recognition application in which feature extraction is the most important step that conditions the efficiency and credibility of all recognition results. In fact, this step relies mainly on hand-crafted techniques which are then categorized by classical classifiers called Machine Learning (ML) algorithms, such as Support Vector Machine (SVM) and Random Forest Trees (RFT) classifiers. Unfortunately, many studies have shown that the majority of the handcrafted methods have reached their limits, especially in applications processing a large number of images. In

general, these methods can provide feature vectors with a very high correlation rate between classes (inter-class) and a shallow correlation rate within the class (intra-class), which leads to serious classification errors. To this end, many efforts have been made to access techniques capable of penetrating deep into the image in search of the best representative features. These features, that should necessarily increase the intra class correlation and reduce the interclass correlation, are called deep features and are extracted by deep methods. These new methods allow images to be analyzed at multiple levels to extract deep features that can increase the feature vectors accuracy, which in turn increases the pattern recognition system accuracy.

The research presented in this dissertation is part of the general context of Computer Aided Medical Diagnosis (CAD) and more particularly in the context of osteoporosis detection. In this work, we propose to use deep features, characterized by their lightness and flexibility. Our method, called Principal Component Analysis Network based osteoporosis detection (OS-PCANet), extracts deep features for high and accurate performance. In addition to the way the PCANet is applied, the proposed method (OS-PCANet) differs from the original method (PCANet) by the way of creating the feature vector using the Histogram of Oriented Gradients (HOG) technique instead of a block-wise histogram. In addition, the selection of dominant features was used to reduce the correlation between different feature vectors. In order to evaluate the proposed strategy, we used an available and well-known database composed of 174 osteoporosis images (87 images from osteoporotic patients (with fracture) and 87 images from healthy people (without fracture).

The dissertation is divided into three chapters and organized as follow:

In the **first chapter**, an overview of the bone structure, its types and the causes of osteoporosis are presented. We conclude this chapter with the diagnostic method using imaging techniques.

The **second chapter** is divided into two main parts, the first part includes the popular techniques used in the feature extraction step, while the second part presents our proposed scheme for the CAD based osteoporosis detection.

The **third chapter** gives the experimental results of the proposed scheme with all the necessary analyzes and discussions, using an available and well-known database of 174 osteoporosis images.

Finally, a general conclusion with future perspectives that we will consider are given at the end of this dissertation.

Chapter 1

Osteoporosis
Risks and Diagnosis

Abstract

Early diagnosis of osteoporosis is an important process in preventing bone fractures. Unfortunately, the lack of simple and practical screening tests is a serious problem as current methods are not suitable for this purpose. In a CAD-based osteoporosis detection system, DXA images are captured and analyzed to extract distinct features for use in the automatic osteoporosis detection system. In this chapter, an overview of the bone structure, types and causes of osteoporosis will be given.

Osteoporosis

Risks and Diagnosis

Osteoporosis is a prevailing bone disease, which weakens the bone and is one of the major factors of disability, especially in elderly persons. Osteoporosis is generally characterized by low bone mineral density mass and micro-architectural deterioration of bone tissue. It is a health problem that more than 200 million people around the world struggle with. In addition, osteoporosis is secondary to heart disease as a global health problem according to the World Health Organization (WHO). In this chapter, an overview of bone structure, their types and causes of osteoporosis will be given.

I.1 Bones

Bones are living functional tissues, which are continually being restructured and have many functions. They structurally support the body, allow us to move and protect our vital organs. They also provide a bone marrow environment in which blood cells are formed and serve as a storage area for minerals, especially calcium. We have around 270 soft bones at birth, some of which fuse as we grow older. As a result, we will have 206 bones as we get into adulthood. The femur is the largest bone in the human body, and the middle ear stirrup is only three millimeters long [1].

I.1.1 Bone structure

Despite the inert appearance of bone, it is a highly dynamic organ that is continuously resorbed and then re-formed. Basically, bones are made up of two main types of tissue, as shown in (Fig. I.1):

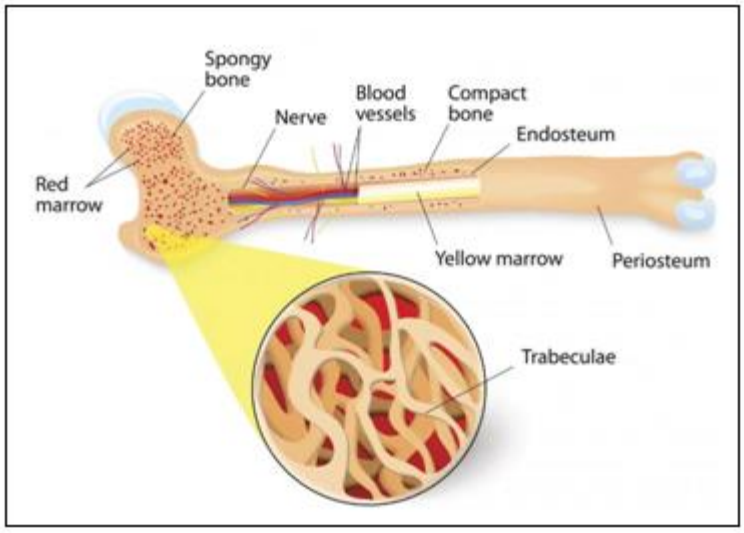


Fig. I.1. Bone anatomy [4]

✍ **Compact (cortical) bone:** A thick, solid and resilient hard-outer layer. It makes up about 80 percent of the bone mass of the adult.

✍ **Cancellous (spongy) bone:** This type of bone contains a network of rod-like structures. It is thinner than compact bone, denser and more durable.

I.1.2 Bone types

Bones have different shapes related to their particular function in the human body, such as protecting and supporting weight. These shapes can be classified into five types:

✍ **Long bones:** They are mainly bones compacted with a little marrow and which would include the majority of the limb bones. These bones tend to support weight and help movement [1].

Long bones, including the femur as well as small bones (inside the fingers), are generally located in the appendicular skeleton and include the bones of the lower limbs (the tibia, fibula, femur, metatarsals, and phalanges) and upper limb bones (humerus, sweep, ulna, metacarpals and phalanges) (see Fig. I.2) [2].

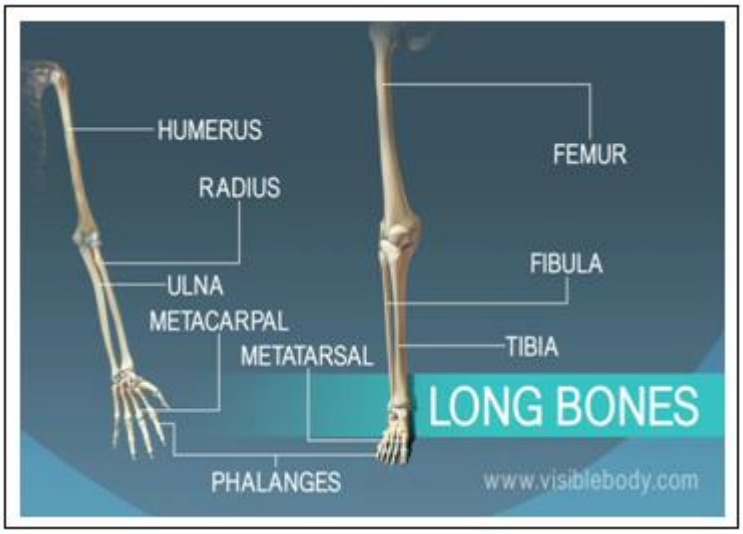


Fig. I.2. Long bones

✍ **Short bones:** Short bones are basely located in the wrist and ankle joints because it is a thin layer of compact bone (see Fig. I.3).

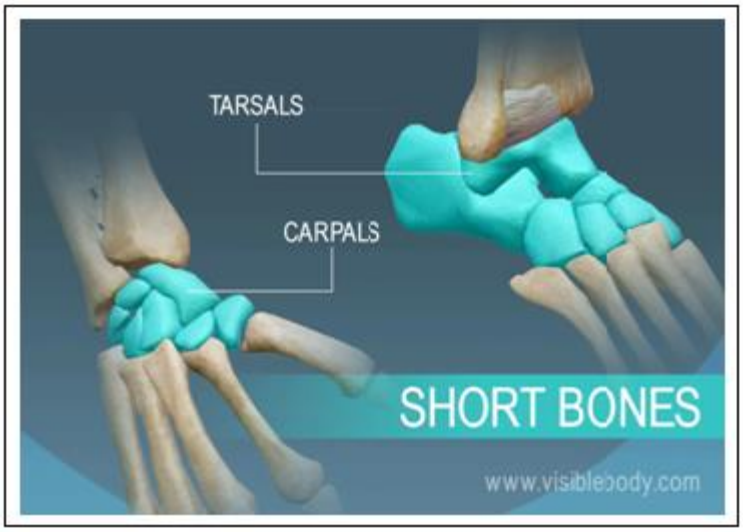


Fig. I.3. Short bones

These consolidated bones of the wrist (carpals) and lower leg (tarsals) provide solidness and smooth movements [2].

✍ **Flat bones:** Flat bones are lean and bend bones. These bones are consisting of two external layers of cortical bone and an inner layer of spongy bone [1].

The main role of these bones is to protect internal organs such as the brain, heart and pelvic organs [2] from shocks. Flat bones can be found in the skull (cranial bones), thoracic cage (sternum or breastbone and ribs), scapulae (wing or shoulder bones), and pelvic bones (see Fig. I.4).

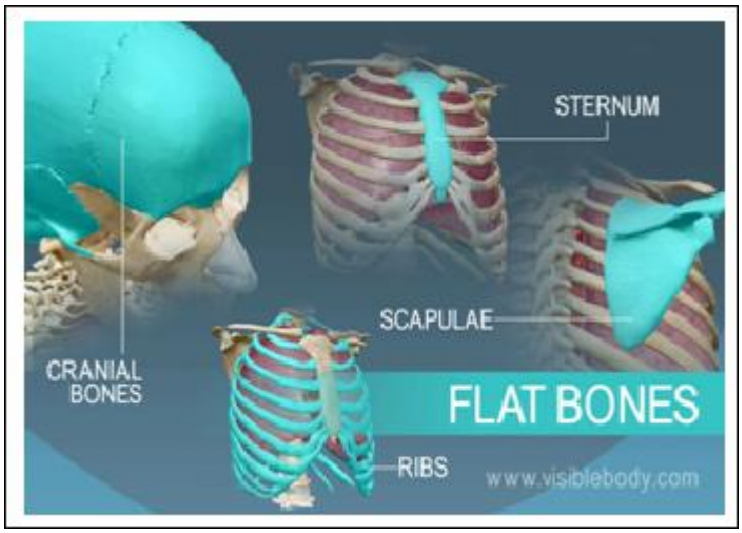


Fig. I.4. Flat bones

✍ **Sesamoid bones:** These small bones (see Fig. I.5). are normally found in the tendons of the hands, knees and feet.

Fig. I.5. Sesamoid bones

The purpose of sesamoid bones is to protect the tendons from stress and wear (for example, the patella which also called the kneecap) [2].

✍ **Irregular bones:** Due to the complicated structure of these bones, they do not fit into any other category.

Irregular bones protect internal organs, such as the vertebrae, which protect the spinal cord, and the pelvic bones (pubis, ilium and ischium) protect the organs of the pelvic cavity (see Fig. I.6).

Fig. I.6. Irregular bones

I.1.3 Bone remodeling

The bone remodeling mechanism consists of the removal of a bone quantum by the osteoclasts, then the formation of new bone by the osteoblasts in the formed cavity. The remodeling process by which bone renews itself occurs constantly throughout life at any one time. When about 10 percent of the bone surfaces of the adult skeleton undergo active remodeling, the remaining 90 percent is found to be quiescent. The length of the remodeling cycle is approximately 6 months, most of this time being occupied by formation; nearly 10 percent of the skeleton is renewed by remodeling each year [3].

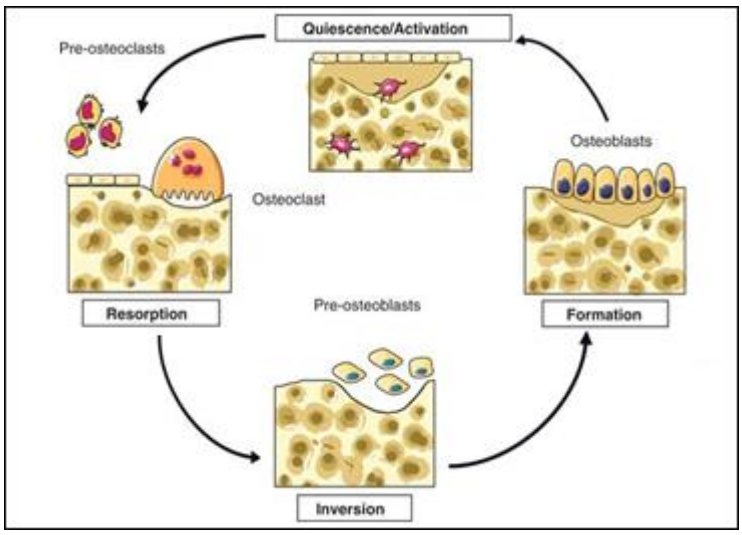


Fig. I.7. Bone remodeling cycle [11]

The remodeling process takes place at the level of the “bone remodeling unit”, which groups the different cell types [4] [5] into four distinct phases (as shown in Fig. I. 7): quiescence/activation, resorption, reversal and formation.

Remodeling allows the body to repair damaged sections, reshape the skeleton during growth, and regulate calcium levels. If one portion of the skeleton, for example during sports or training, is placed under high pressure over time, the areas of the bone under most pressures can become denser. Several bodily hormones, including parathyroid hormone, calcitonin, vitamin D, estrogen in women, and testosterone in men, controlling remodeling process [6][1].

I.2 Osteoporosis

Osteoporosis occurs when the creation of new bone does not follow the loss of the old bone. It makes bones weak and brittle and can cause fractures.

I.2.1 Definition

Osteoporosis is defined as a skeletal disorder characterized by the weak bone strength, which cause to an increased risk of bone fracture to a person. The word osteoporosis literally means "porous bone" [7], this happens when the bones lose too much of their mineral content, especially calcium. Bone density is reduced over time, and therefore their strength, in which they become brittle and break quickly, for example, a person with severe osteoporosis may break their bones with a simple sneeze or sudden movement. The chances of developing osteoporosis depend on the density of a person's bones earlier in life. Diet, health and physical exercise will also determine bone density throughout life. (Fig. I. 8) clearly shows the bone density of healthy bone and bone with osteoporosis [8].

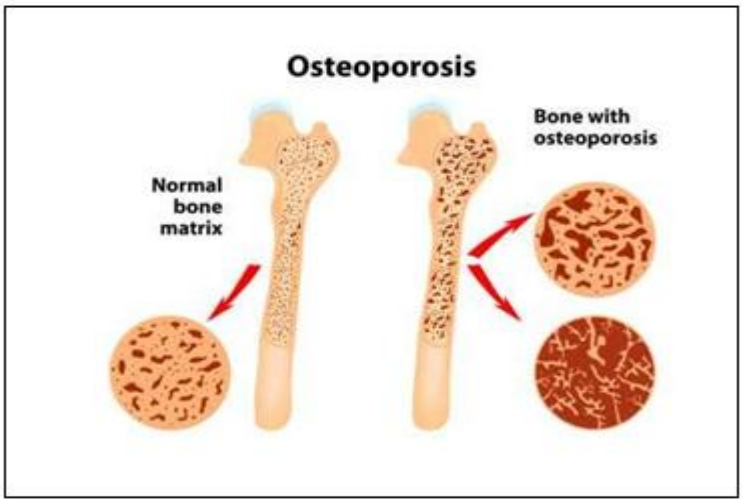


Fig. I.8. Osteoporosis disease

Bone is constantly renewed in a two-step process (resorption and formation) that occurs throughout life. With age, bone loss slowly begins to increase. With increased bone loss, the rate of its generation will also decrease, which means that during childhood more bone is

produced than is removed, reaching its maximum mass and strength in the mid-30s. Subsequently, bone formation becomes slower than resorption, leading to a slow decline in bone density in the skeleton [8] [9].

Bone fractures are usually the first symptom of the osteoporosis disease, which can damage any bone. Fractures of the hip, spinal cord and wrist are the most common, which is of particular concern due to hospitalization and often requiring surgery, which can have serious consequences such as chronic disability and even death [10].

I.2.2 Osteoporosis Types

There are four different types of osteoporosis, as described below:

✍ **Primary osteoporosis:** Primary osteoporosis is the most common type of osteoporosis, affecting more women than men; it is usually caused by factors related to age. Peak bone density (mass) is reached between the ages of 25 and 30. (Fig. I. 9) illustrates the decrease in bone mass with age.

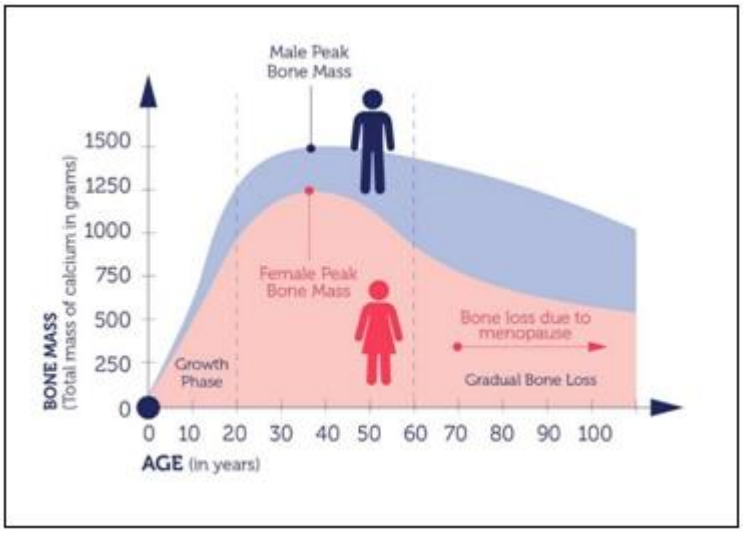


Fig. I.9. Decreasing bone mass with age

The accelerated loss of bone density will usually begin after a woman's menstrual cycle has ended, which happens when estrogen production begins to slow (this is usually around the age of 45 to 55). In men, gradual bone loss will normally begin between the ages of 45 and 50; This is when the testosterone production begins to slow down. Usually, it only influences people over the age of 60 [9] [10].

✍ **Secondary osteoporosis:** This type of osteoporosis can affect anyone at any age; however, secondary osteoporosis occurs as a result of certain medical conditions such as leukemia, hyperthyroidism and can also result from taking certain medications that cause loss of bone density, including high doses of thyroid hormone replacements or drugs known aromatase inhibitors (which are used in the treatment of breast cancer) [9].

✍ **Osteogenesis imperfect:** This type of osteoporosis is extremely rare and leads to bone fracture without a logical explanation. Osteogenesis imperfecta is observed at birth.

✍ **Idiopathic juvenile osteoporosis**: This type of osteoporosis is also very rare and most often occurs during the rapidly growing period of a child's life (children between 8 and 14 years old). Idiopathic juvenile osteoporosis usually affects children with a history of being overweight before puberty and increases the risk of bone fractures. The cause of this type of the disease is not yet known and there is no cure [9] [10].

I.3 Osteoporosis development: risks and causes

Osteoporosis is called the silent disease because bone loss occurs without symptoms. Often people don't know they have the disease until a bone break, frequently in a minor fall that wouldn't normally cause a fracture. Compression fractures of the spine are a common occurrence, in which they can occur even after seemingly normal activity, such as bending or twisting to pick up a light object. Fractures can cause severe back pain, but sometimes go unnoticed, the vertebrae collapse on themselves and the person who actually loses height. The hunchbacked appearance of many older women, sometimes are called (dowager’s) hump or (widow’s) hump, is due to this effect of osteoporosis on the vertebrae [11] [12].

There are several risk factors that can cause this disease, such as age, gender, lifestyle, etc., as shown in the following table:

Table I.1. Osteoporosis risk factors

Factors	Description
Age	Osteoporosis is more likely as people get older and their bones lose tissue.
Gender	Women start with less bone, and they also lose bone tissue more quickly with age. While women typically lose 30 to 50 percent of their bone mass over their lifetimes, men only lose 20 to 33 percent.
Race	Caucasian and Asian women are most at risk of contracting the disease, but African American and Hispanic women can also contract it.
Figure type	Women with small bones and those who are thin are more liable to have osteoporosis.
Early menopause	Women who stop menstruating early because of heredity, surgery, or a lot of physical exercise can lose large amounts of bone tissue early in life. Conditions such as anorexia and bulimia can also lead to early menopause and osteoporosis.
Family history	Parental history of hip fracture.
Lifestyle	People who smoke or drink too much or who don't exercise enough have an increased risk of osteoporosis.
Diet	Those who do not get enough calcium or protein are more likely to suffer from osteoporosis. That is why people who follow a constant diet are more prone to the disease.
Genetics	Research in Europe reported in 2003 that variations in a gene on chromosome 20 could make some postmenopausal women more likely to develop osteoporosis. Studies were continuing on how to identify the gene and use the information from research to prevent osteoporosis in carriers [12].

I.4 Osteoporosis Diagnosis Based on Imaging Techniques

Bone Mineral Density (BMD) is a measure of the amount of minerals (mainly calcium and phosphorus) contained in a certain volume of bone. These assessments are used to diagnose osteoporosis, assess the effectiveness of osteoporosis therapies, and predict the likelihood of bone fractures.

I.4.1 Imaging techniques

There are a number of X-ray tools available to measure bone mineral content and the fracture risk. Indeed, Dual Energy X-ray Absorptiometry (DXA), Computed Tomography scans and Magnetic Resonance Imaging (MRI) [13] are an example of these tools.

✍ **X-Rays:** The test is based on the fact that different parts of the body attenuate (stop) X-rays better than others. X-rays are ionizing radiation, generated by an X-ray tube. The rays are controlled by shielding a narrow beam, directed towards the part of the body being examined. On the opposite side of the body, an X-rays film is positioned in the X-rays path, and is "exposed".

The X-ray film is then processed and an image generated, such as the image plotted in (Fig. I. 10.) The part of the body easily crossed by X-rays gives part of the film black.

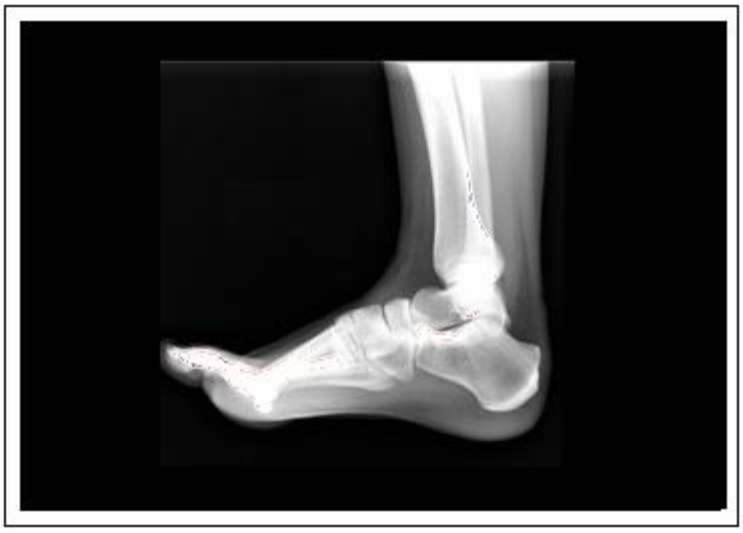


Fig. I.10. X-ray of feet right side

Where the X-rays are stopped, the film will be white and different shades of gray in between. X-ray exams are used to examine many structures in the body. They are particularly good for looking at the chest and limbs, but a plain X-ray is a very versatile imaging method.

✍ **Dual-Energy X-ray Absorptiometry:** Dual-energy X-ray absorptiometry (DXA or DEXA) is an imaging technique developed in the mid-1980s and approved for the bone mineral density (BMD) measurement in clinical practice by the Food and Drug Administration (FDA) in 1988. DXA scanners were originally designed to measure mineral content and bone density [14] (see Fig. I. 11).

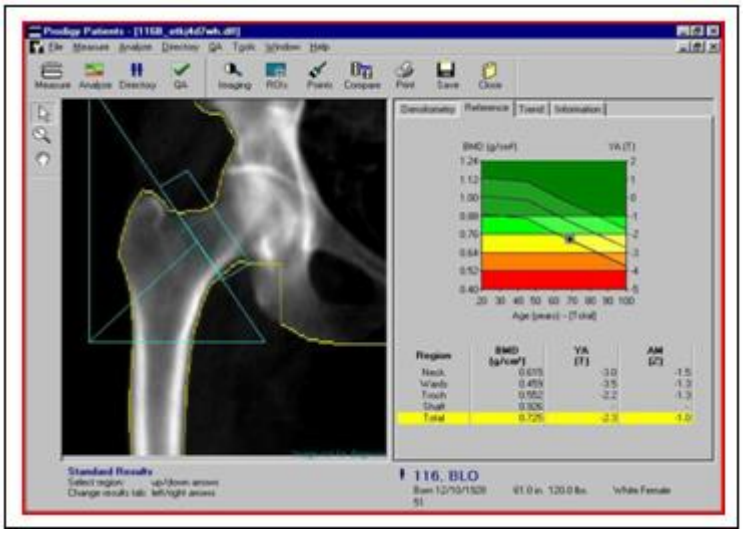


Fig. I.11. DEXA bone density scans

It is considered the gold standard non-invasive method for the diagnosis of osteoporosis, for the assessment of fracture risk and for the monitoring of BMD. It is obtained by passing filtered X-ray beams at two photon energies throughout the body which are differentially attenuated by the material in their path [15].

In the DXA test (see Fig. I. 12), two X-rays of different energies irradiate the bone in order to discriminate the contribution of two types of tissue, namely hydroxyapatite (representative of bone tissue) and soft tissue. Further elaboration of the radiation transmitted by the bone sample and revealed by the detector provides a 2D grayscale image, representing a BMD map over the fully irradiated area.

Fig. I.12. DXA scan machine

An edge detection algorithm is then used to find the bone edges. After subtracting the soft tissue contribution, the average area BMD (g/cm2) is calculated by averaging the BMD

values of all the pixels included in the edge contour, obtaining a measure of the apparent density of the bone [14] [15] which are then converted to a T-score or a Z-score (see Fig. I. 13) using the following formula [13]:

$$T = \frac{patient's\ BMD - population\ BMD}{standard\ deviation\ of\ the\ population\ peak\ BMD} \quad (1)$$

$$Z = \frac{patient's\ BMD - population\ age\ related\ BMD}{standard\ deviation\ of\ the\ population\ age-related\ BMD} \quad (2)$$

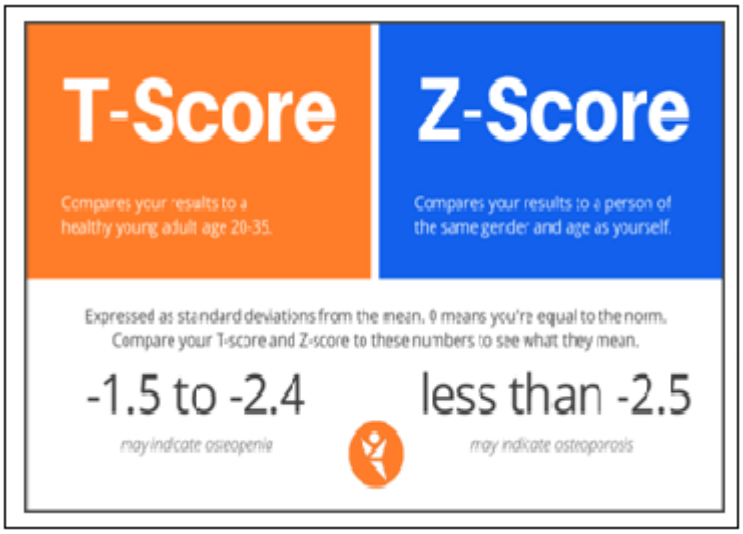


Fig. I.13. Bone density test results

The scores presented from the bone density test results consist of two numbers: A T-score and a Z-score, both expressed as standard deviations from the mean, or average. The T-score is a comparison of results with a healthy young adult 20 to 35 years old, while the Z-score is a comparison between you and someone of the same gender and age as yourself. The more standard deviations less than 0, expressed as negative numbers, the lower your BMD and the higher your risk of fracture. A score of 0 indicates that you are equal to the norm. The implications of osteopenia arise for scores of -1.5 to -2.4 and a score of -2.5 or less indicates that it is osteoporosis.

✍ **Magnetic Resonance Imaging Techniques:** Magnetic Resonance Imaging (MRI) is a non-ionizing technique that provides indirect information on the 3D organization of bone by measuring the signal from hydrogen atoms contained in water and fatty compartments of the bone marrow bone. It is a method of looking inside the body without the need for surgery or x-rays. The MR scanner is a large doughnut shaped magnet open at both ends (see Fig. I. 14).

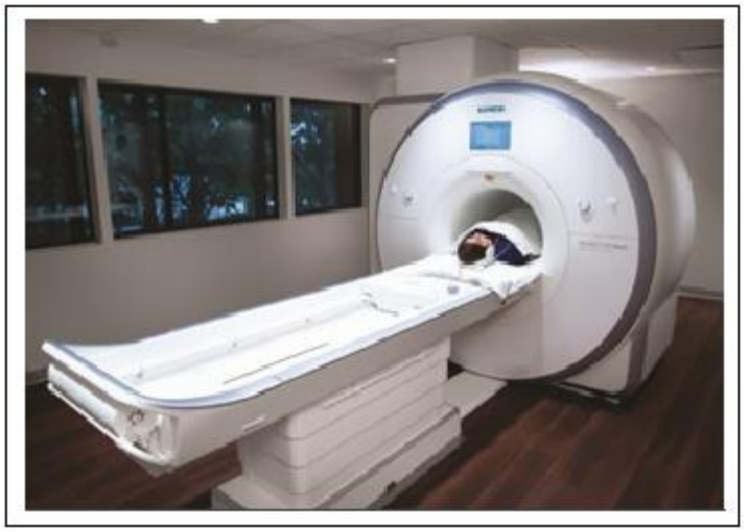


Fig. I.14. MRI scanner

It uses a strong magnetic field, radio waves, and a computer to produce clear images of the human body. This technology is important because MRIs can show doctor the difference between healthy tissue and diseased tissue (see Fig. I. 15).

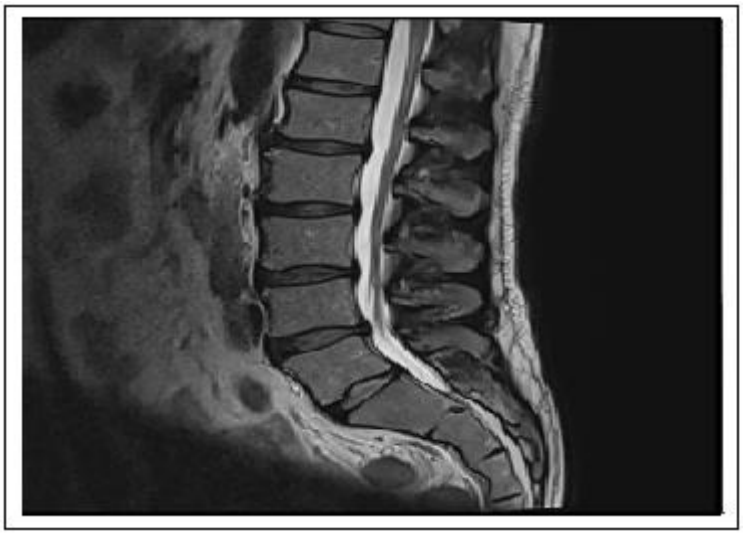


Fig. I.15. MRI image for lumbar spine

However, the low amount of water contained in bone and the presence of surrounding tissue results in a low signal-to-noise ratio (SNR) in MRI images. For this reason, MRI is typically used at peripheral sites such as the distal radius, distal tibia, and calcaneus. This method has shown promising results in the evaluation of bone density, but has certain limitations for clinical use, due to the long imaging times (approximately 1 h) and the strong magnetic field applied [13]. In general, MRI techniques require expensive equipment and a complicated post-processing phase, which reduces their availability as routine OP diagnostic tools in clinical applications.

✍ **Computed Tomographic Imaging Techniques:** Computed tomography (CT), (see Fig. I. 16), is a diagnostic imaging test used to create detailed images of internal organs, bones, soft tissues and blood vessels. Sectional images generated during a CT scan can be reformatted in multiple planes and can even generate three-dimensional images that can be viewed on a computer screen or transferred to electronic media. This imaging technique makes it possible to assess bone organization in 3D by irradiating the bone with an X-ray beam in different directions and by detecting the transmitted radiation. Since the attenuation of the x-ray beam depends on the density of the material passed through, the revealed radiation intensity profile contains the compositional information of the bone structure along the direction analyzed. Successively, these images are produced to perform morphometric and densitometric analyzes on the scanned bone sample.

Fig. I.16. CT imaging machine

At present, there is a big difference between the performance shown by CT instruments used on patients in clinic and that of laboratory CT scans used to analyze the 3D organization of the bone structure of bone vivo samples. Clinical CTs are indeed characterized by low resolutions, of the order of a few hundred μm, while laboratory CTs can achieve resolutions of less than 1 μm [13] [14].

I.4.2 Diagnosis

Some types of doctors may have more training and experience than others in diagnosing and treating people with osteoporosis. These include a geriatrician, specializing in the treatment of the elderly; an endocrinologist, specializing in the treatment of diseases of the body's endocrine system (glands and hormones); and an orthopedic surgeon, who treats fractures such as those caused by osteoporosis.

Before making a diagnosis of osteoporosis, the doctor usually takes a complete medical history, performs a physical exam, and orders x-rays, as well as blood and urine tests, to rule out other conditions that lead to loss of bone mass. The doctor may also recommend a bone density test. This is the only way to know if osteoporosis is present. It can also show how far the disease has progressed.

Current clinical practice guidelines for the diagnosis and management of osteoporosis suggest that individuals, men and women, over the age of 50 be evaluated for risk factors associated with osteoporosis and factors associated with high risk of fracturing (see Fig. I. 17). Some risk factors include measuring height annually, assessing for the presence of a vertebral fracture, assessing a history of falls within the past year, and performing blood tests [17]. However, the most accurate way to diagnose osteoporosis is to use an instrument to measure bone density [13].

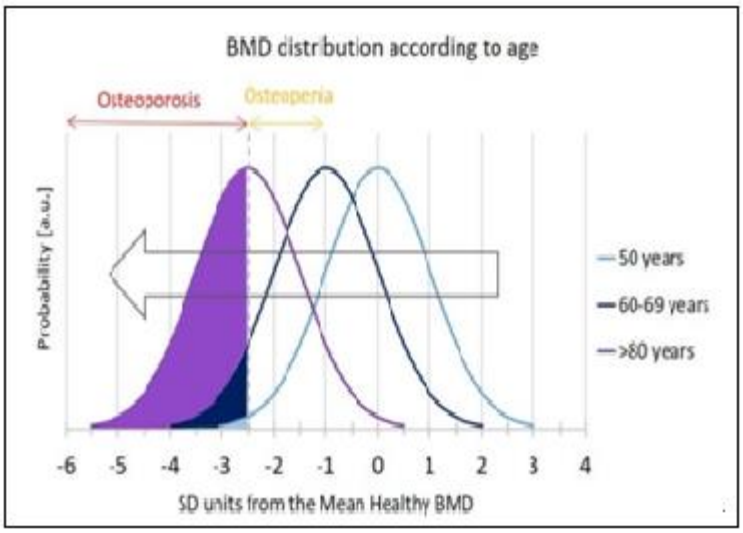


Fig. I.17. BMD distribution according to age [23]

Radiography works by emitting X-ray beams that pass through the bones and detecting what is not absorbed. The denser the bone, the more X-ray energy is absorbed, while less dense and more porous areas absorb less energy. A "picture element", the radiant energy per pixel, is created and then converted to an "areal density value", measured in grams/centimeter (g/cm) [18].

The problem with simple radiography is that it produces low-sensitivity images that may not accurately diagnose osteoporosis. Likewise, the DXA works by emitting two X-ray beams of different energies (dual X-ray absorptiometry) and the values can be expressed in g/cm2, the DXA is currently considered as the gold standard for the diagnosis of osteoporosis and

the accuracy of DXA at the hip exceeds 90 percent, it supports the assessment of fracture risk in older women.

Although MRI techniques and CT scans can be used to diagnose osteoporosis, MRI requires expensive equipment and a complicated post-processing phase, which reduces their availability as routine osteoporosis diagnostic tools, while the CT scanner must be calibrated to convert the results into units relevant to BMD, which makes this technique also not very useful in routine diagnosis due to high radiation exposure, high cost and quality control difficulties compared to DXA [12].

For women, there are four general diagnostic categories proposed by WHO for assessments performed with DXA (see Table I.2):

- **Normal:** Hip BMD T-score greater than -1 SD lowers than the young adult female reference mean (T-score ≥ -1.00 SD).
- **Osteopenia:** Hip BMD T-score less than -1.00 SD lower than the young adult female reference mean but greater than -2.5 SD.
- **Osteoporosis:** Hip BMD T-score less than -2.5 SD lower than the young adult female reference mean.
- **Severe osteoporosis**: Hip BMD T-score less than -2.5 SD lower than the young adult female reference mean and the presence of one or more fragility fractures.

Table I.2. WHO osteoporosis classification

Diagnosis	T-score
Normal	>-1.0
Osteopenia	<-1.0,>-2.5
Osteoporosis	<-2.5
Severe osteoporosis	<-2.5 plus fragility fractures

Cutoff values for men are less defined than for women, but a number of studies have indicated that similar cutoff values for hip BMD in women can be used to diagnose osteoporosis in men [12].

A new products and methods have been developed through molecular and cellular research focusing on bone fragility; it has become essential to develop efficient and sensitive non-invasive means for detecting early changes in the fracture repair process [18]. New specialized non-invasive and/or non-destructive techniques capable of providing structural

information on local and systemic skeletal health, fracture propensity and the pathophysiology of bone fragility have been developed, allowing quantitative assessments of bone macro and microstructural features, and improve our ability to estimate bone strength. DXA and CT, especially volumetric QCT (vQCT), can provide a quantitative assessment of bone macrostructure, while high-resolution CT (hrCT), microCT, high-resolution MR (hrMR), and microMR can achieve an assessment of the trabecular bone microstructure (see Table I.3) [17].

Table I.3. Overview of CT techniques to determine BMD and bone architecture

	vQCT	**hrCT**	**microCT**
In plane pixel size	>0.3×0.3 mm^2	0.1×0.1-0.3×0.3 mm^2	Isotropic
Slice thickness	>1 *mm*	0.2-1 *mm*	1-100 μm
Equipment	Whole-body clinical scanners; dedicated peripheral scanners.	Whole-body clinical scanners; dedicated peripheral scanners.	Dedicated micro CT scanners.
Skeletal location	Spine, hip, forearm, tibia.	Spine, forearm.	Human biopsies: iliac crest, animals and specimens: various.
Subjects/samples	Human in vivo.	Human in vivo/human biopsies/bone specimen.	Laboratory animals in vivo and in vitro, bone specimen.
Applications	BMD/bone macrostructure/FEM.	Bone macrostructure/trabecular microstructure.	Trabecular and cortical microstructure/micro FEM.

vQCT, hrCT and hrMR are generally applicable in vivo, while microCT and microMR are mainly used in vitro [8]. These advanced imaging modalities currently available make it possible to investigate on bone fragility and to define the skeletal response to innovative therapies and to evaluate the biomechanical relationships [18]. QCT allows separate analysis

of the trabecular and cortical compartments. Analysis of cortical bone, especially at the hip, is important in estimating the fracture risk, and this technique has been used in several clinical trials [19].

I.5 Conclusion

In this chapter, we have defined and described osteoporotic disease and its different diagnostic techniques. We first talked about bone and its types, then we talked about the process of bone remodeling in the human body, and then osteoporotic disease and its types. Finally, on its diagnosis and imaging techniques for its evaluation and the factors that increase the risk of developing this disease.

Chapter 2

Discriminative Features Extraction

Need and methods

Abstract

The majority of research in the field of pattern recognition is aimed at developing new techniques, especially those aimed at the feature extraction task, to transform raw data into other representations that honestly illustrate their content and can then be used to analyze and understand it. In this chapter and before detailing our proposed osteoporosis detection scheme, we will attempt to provide an overview of common feature extraction techniques.

Discriminative Features Extraction

Need and methods

Feature extraction is an important task in the pattern recognition applications due to a large amount of different existing features in signal (especially in image) and his multiple areas of applications that it covers medical imaging, biometrics, remote sensing and robotics. Due to this necessity, a well considerable effort has been made by the researchers in this direction, resulting in many cases an excellent enough classification result. In this chapter and before detailing our proposed osteoporosis detection scheme, we will attempt to provide an overview of common feature extraction techniques.

II.1 Pattern recognition system

Pattern recognition is becoming an important task for the success of a diverse range of applications, including medical diagnostics, image and video retrieval, video surveillance, biometrics and social media networks. The field of medical diagnostics is one of the most important areas that have been significantly affected by this trend, especially with the massive technological advancements in the field of digitization and automation. One of the most challenging tasks in the pattern recognition system is to efficiently extract salient features which are able to representing the discriminating information in the captured signal, since good results are directly related to the uniqueness and variability of those features that are used to distinguish different patterns.

II.1.1 Basic structure

The main objective of the pattern recognition system is to classify features (pattern) into categories (classes) based on observations (features vector) extracted from them. This system

can include a training phase which is used to train the system to recognize a pattern from the data. When this phase is performed, the system will then be able to operate to recognize unknown patterns. A basic pattern recognition system consists of four modules (see Fig. II.1) some of which are common to both the training phase and the classification phase: acquisition, feature extraction, prediction (measure of similarity) and decision. The acquisition and feature extraction takes place during training and classification. However, prediction and decision are only used during the classification phase.

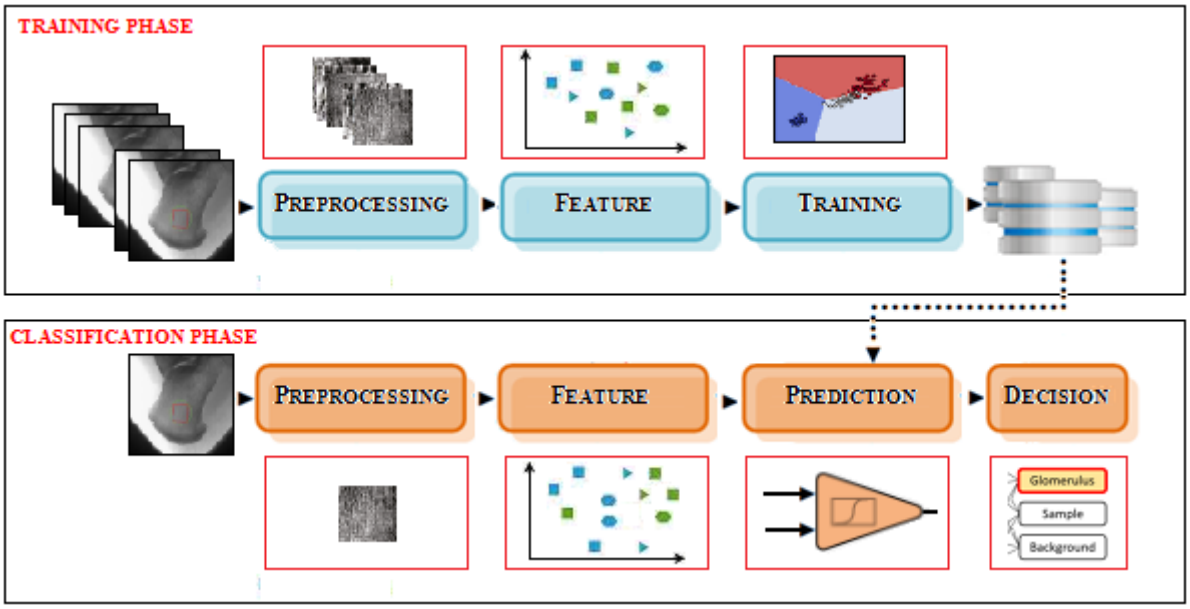


Fig. II.1. Basic structure of the pattern recognition system

✍ **Acquisition:** The acquisition device used depends on the data nature to be analyzed (context work), in which a sensor is necessary to acquire the signal and then convert it into a digital format so that it can be processed by a computer.

Generally, the signals (or images) acquired from the sensors (raw data) cannot be used directly by the systems and they must undergo a **preprocessing** whose role is to configure the raw data in order to prepare them for the feature extraction step. Indeed, signal and image processing tools are the main sources of pre-processing methods, such as filtering, contrast enhancement, segmentation, normalization, etc.

✍ **Feature extraction:** In general, the acquired data contains several features that can be used to recognize them. Indeed, the type of features used is defined according to the intended application. For example, in facial recognition, it is interesting to take into account the different patterns, such as ridges, that can be considered as patterns. Lines can also be used to represent facial expressions. When detecting linear features such as road detection, it is

necessary to use line-based feature extraction methods. Finally, for medical imaging, texture-based feature extraction methods may give better results.

✍ **Prediction:** In this step, the feature vector of an examination input data sample must be simulated, with the model obtained during the training phase, in order to determine its expected class. This operation is important in the field of classification.

✍ **Decision:** In the decision (or classification) step, a label (class) is assigned to the input data sample according to the probability of simulation between the extracted feature vector and the model obtained during the training phase.

II.1.2 Phases of pattern recognition

In general, all pattern recognition systems share the same architecture, which operates in two distinct phases: training (learning) and classification.

✍ **Training:** In training, the system is provided by predefined patterns which will allow the recognition system to be tuned so that it is able to recognize subsequent patterns of unknown class. In general, such a pattern recognition system comprises many parameters which must be adapted. However, this system is designed so that the parameters can be defined during this phase from the training data.

✍ **Classification:** The task of the classification phase is to identify the class of the input data sample. During this phase, the input data sample is analyzed and a set of parameters (feature vector) is extracted as during the training phase. In the classification phase, the sensor used must have properties as close as possible to the sensor used during the training phase.

II.2 Feature extraction

In modern life, vision technology gives machines the ability to replace or accompany manual tasks using multiple vision sensors that used to acquire the data to be processed. Overall, many machine vision applications are based on images that are an essential aspect of a data representation widely used in medical imaging, robotics and computer vision, industrial vision, biometrics, and remote sensing. This representation of data becomes the dominant technology for machine vision applications and is particularly useful when the texture or color of the object is the key to extract distinctively features effectively.

Feature extraction from image is the process of transforming the image’s raw pixel values into more meaningful and useful information that can be used in a machine vision application. Thus, these processes have been carefully designed by specialists (Hand-craft) using their

knowledge and expertise. Thus, with the advent of convolutional deep learning, the neural networks themselves automatically do the work of the expert in the convolutional layers. Generally, all feature extraction methods are classified into two main categories:

II.2.1 Hand-craft techniques

The image contains many features that can be used to analyze such a scene. The type of feature used varies depending on the intended application. For example, in an image containing a scene consisting of three patterns, several methods should be used to extract each pattern separately. Therefore, it is necessary to know the type of scene to effectively choose the appropriate method. As these techniques require expertise, the features obtained are considered to be hand-crafted (feature engineering). A weakness of these methods is that the features of the image are calculated according to a set of predesigned steps and do not depend on the image context. In the following subsection, an overview of some hand-crafted feature extraction methods, which were used in this work, is carried out.

✍ **Bag of visual words (BoVW):** The BoVW comes from a technique called Bag of Words (BoW), which is a feature extraction method with text data, this approach is a simple and flexible way to extract features from documents. A bag of words is a text representation that describes the occurrence of words in a document, and just keeps track of the word count and ignores grammatical details and word order. This is called a "bag" of words because any information about the order or structure of words in the document is removed. The model only deals with whether known words appear in the document, not where in the document. Similarly, in the computer vision field, the same concept is used in the bag of visual words. Here, instead of taking the word from the text, the image patches and their feature vectors are extracted from the image in a bag.

In BoVW, the interest points or keypoints (the unique points in an image, and even if the image is rotated, shrink, or expand, its keypoints will always be the same) of the images are detected and described. Then, the keypoints are grouped according to their descriptors (the descriptor is a description of the keypoints), then the *k* most representative samples are selected for the composition of the vocabulary of visual words (Codebook). Once the visual word vocabulary has been constructed, a visual word histogram is generated for each image. These histograms are used as feature vectors for image classification [20] [21].

After building the vocabulary, each image in the dataset is represented by visual word histograms which are later used as a feature vector for the classification of images. Histograms are created based on the frequencies of visual words that compose the vocabulary

of images. Therefore, the visual word with the shortest distance will have its histogram's bin incremented by one unit. The most commonly used distance function, which computes the distance between a keypoint and the visual words is the Euclidean distance.

Feature detection and description: Feature detection is the identification of particular local features in the image (corners, edges, points of interest), such as: Harris, Shi-Tomasi and FAST keypoint detectors. The main property of local feature detection algorithms is repeatability, that is, given two images of the same object or the same scene, taken under different viewing conditions, a high percentage of the features detected on the scene part visible in both images must be found in both images. Besides the repeatability property, detectors with good features should have distinctiveness, locality, quantity, accuracy and efficiency. The importance of these different properties depends on the application and the actual parameters. Once a set of local features has been detected from an image, some measurements are taken from a region centered on the local features and converted to local descriptors. The main task of a keypoint descriptor is to describe an interesting patch in an image. There are a variety of local descriptors to describe the content of the image such as SIFT, SURF, ORB [20]. Later, we summarize the SIFT and ORB algorithms, which are the most commonly used local detection/description algorithms for visual recognition tasks.

Dictionary (Codebook): The visual word dictionary that can be used to define each image in terms of the frequency of each word present in an image is known as the bag of visual words. Clustering is the assignment of objects to groups of similar objects (clusters), (see Fig. II.2). Objects are generally described as feature descriptors (also called attributes). A distance measure is a function that quantifies the similarity of two objects [22].

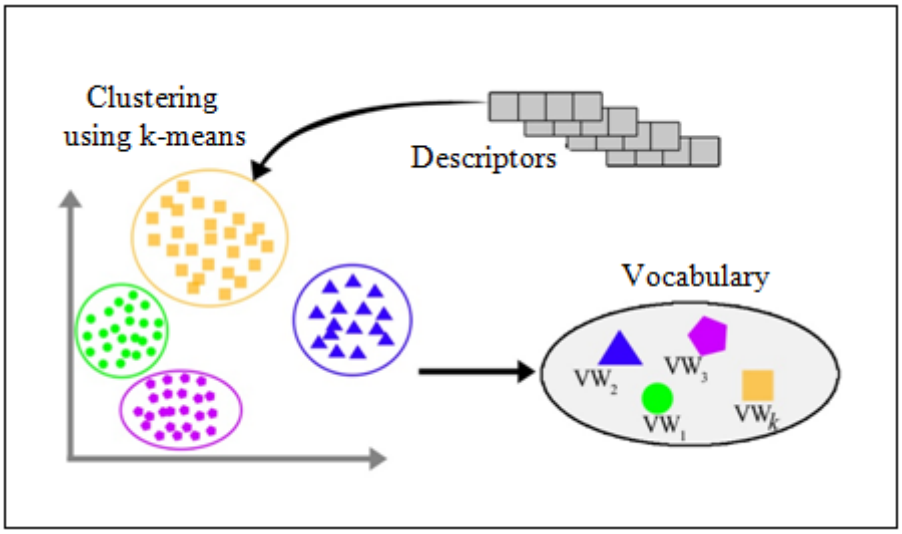


Fig. II.2. *k*-means clustering

The most widely used algorithm for clustering is k-means, which is an unsupervised algorithm [23]. (Fig. II. 3) shows a bag of visual words model showing the images as a visual word histogram. Now this gives a fixed length feature vector.

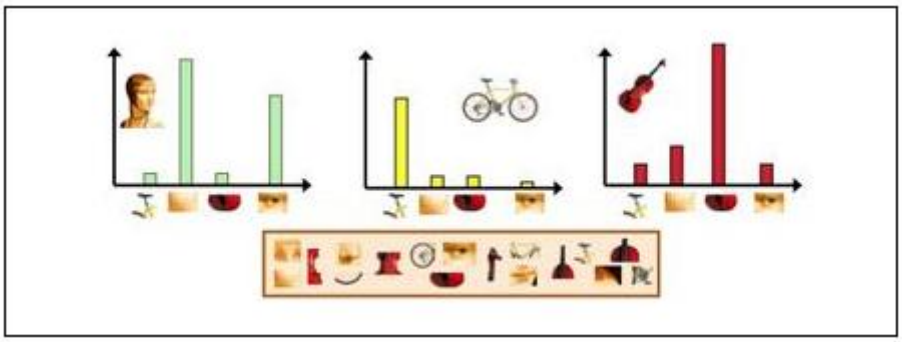


Fig. II.3. Bag of Visual Words model

✍ **Scale Invariant Feature Transformation (SIFT):** SIFT is invariant at the scale and at the rotation of the image. This is an algorithm that describes features of the image which have many properties in order to make them suitable for matching different images of an object or scene. The features are invariant to the scaling and rotation of the image, and partially invariant to the change in lighting and from the 3D camera viewpoint. They are well localized in the spatial and frequency domains, which reduces the probability of disruption by occlusion, clutter or noise. In addition, the features are very distinctive, which allows a single feature to be correctly compared with high probability to a large feature in database, providing a basis for object and scene recognition. The cost of extracting these features is minimized by taking a cascade filtering approach, where the most expensive operations are applied only to locations that pass an initial test. The major calculation stages used to generate all the image features are described below:

- ***Scale-space extrema detection***: The first stage of computation searches over all scales and image locations. It is efficiently implemented by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation.

- ***Keypoint localization:*** At each candidate location, a detailed model is fitted to determine the location and scale. Key points are selected based on measurements of their stability.

- ***Orientation assignment:*** One or more orientations are assigned to each keypoint location based on the local image gradient directions. All future operations are performed on image data that has been transformed with respect to the orientation, scale, and location assigned to each feature, thus providing invariance to those transformations.

- ***Keypoint descriptor:*** Local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows significant levels of local shape distortion and change in illumination [24].

SIFT is the most widely used local approach for recognition tasks. It was originally proposed as combination of a difference-of-Gaussian (DoG) interest region detector and a histogram of gradient (HoG) locations and orientations feature descriptor. The SIFT descriptor is invariant to scale, rotation, affine transformations and partially invariant at illumination changes [20].

✍ **Oriented FAST and Rotated BRIEF (ORB):** ORB (Oriented FAST and rotated BRIEF) is a fusion of the FAST keypoint detector and the BRIEF descriptor with some additional features to improve performance. ORB is a very fast binary descriptor based on BRIEF, which is rotation invariant and noise resistant. ORB is proving its effectiveness in several real-world applications, including object detection and patch tracking on a smartphone. ORB is composed of the well-known FAST keypoint detector and the BRIEF descriptor; for this reason, it is called ORB (Oriented FAST and Rotated BRIEF). These techniques offer good performance and have a low cost. A fast and accurate orientation component is added to FAST, along with efficient computation of oriented BRIEF features, analysis of variance and correlation of oriented BRIEF features, and a learning method to decorrelate BRIEF features under invariance rotational, leading to better performance [24].

- ***Features from Accelerated Segment Test (FAST):*** FAST is a corner detection method, which could be used to extract feature points and later used to track and map objects in many computer visions tasks. The most promising advantage of the FAST Corner detector is its computational efficiency. In addition, when machine learning techniques are applied, higher performance in terms of computing time and resources can be obtained. The FAST corner detector is very suitable for real-time visual recognition applications due to its high-speed performance [26].

- ***BRIEF*** is a feature descriptor that uses simple binary tests between pixels in a smoothed image patch. Its performance is similar to SIFT in many ways, especially in terms of robustness to lighting, blurring and perspective distortion. However, it is very sensitive to in-plane rotation [27].

✍ **Gray Level Co-occurrence Matrix (GLCM):** A Gray Level Co-occurrence Matrix (GLCM) is a well-known approach to extract second-order statistical texture features using statistical distributions of intensities values combinations at different positions relative to each other in an image. Depending on the number of intensity points in an image, the statistics are divided into first, second, and higher order. Higher order statistics are theoretically possible but cannot be implemented due to the computation complexity. Texture features contain information about the structural arrangement of surfaces and their relationship to the surroundings. In total, twenty-two texture-based features are obtained, including energy, correlation, entropy, inverse difference moment (IDM), homogeneity, sum variance, autocorrelation, contrast, maximum probability, dissimilarity, normalized IDM and many more. GLCM is just a feature detector which is a technique that looks for the relationship of intensity in images. He builds a new matrix which has different relations of intensity. The idea behind GLCM is to describe texture as a matrix of "pair gray level probabilities" [27-29].

- First-order texture measurements are statistics calculated from values in the original image, such as variance, and do not take into account relationships between pixels.

- The second order means they consider the relationship between the groups of two pixels in the original image. Note that, this is not the same as "second order equations", which would mean equations with certain variables squared.

- Third and higher order textures (considering the relationships between three or more pixels) are theoretically possible but not implemented due to computation time and interpretation difficulties [30].

The GLCM is a two-dimensional matrix in which each element pij represents the frequency of occurrence of a pair of pixels (where i and j are the gray levels) in a spatial relationship separated by distance and angle α. Let G be an image texture of size $M \times N$. An element p_{ij} can be calculated by counting the number of relations with the following equation [31]:

$$p_{ij} = \{G(m,n) = i, G(m+\delta, \mathrm{n}+\delta) = \mathrm{j}\} \text{ for each } \alpha, \qquad (1)$$

Where $m = 0,1,\dots,M-1$, $n = 0,1,\dots,N-1$, and $\alpha = 0°,\dots,360°$. Some texture features including entropy (E), contrast (Con), correlation (Cor), energy (Eg) and homogeneity (H) can be given as follows:

✓ ***Entropy:*** Entropy shows the total amount of image information that is required for image compression. Complex textures generally have high entropy. Entropy is strongly, but inversely correlated with energy.

$$E = -\sum_{ij}(p_{ij}\log p_{ij}) \quad (2)$$

✓ ***Contrast:*** Measures neighborhood variations in the gray level co-occurrence matrix and spatial frequency of an image and the difference moment of GLCM. It is the difference between the highest and lowest values of a contiguous group of pixels. Contrast measures the amount of local variation within the image.

$$Con = \sum_{ij}(|i-j|^2 p_{ij}) \quad (3)$$

✓ ***Correlation*****:** It gives the joint probability occurrence of the specified pixel pairs.

$$Cor = \frac{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(ijp_{ij}-\mu_1\mu_2)}{\sigma_1^2\sigma_2^2} \quad (4)$$

Where:

$$\mu_1 = \sum_{i=0}^{M-1} i \sum_{j=0}^{N-1} p_{ij} \quad (5)$$

$$\mu_2 = \sum_{i=0}^{M-1} j \sum_{j=0}^{N-1} p_{ij} \quad (6)$$

$$\sigma_1^2 = \sum_{i=0}^{M-1}(i-\mu_1)^2 \sum_{j=0}^{N-1} p_{ij} \quad (7)$$

$$\sigma_2^2 = \sum_{i=0}^{M-1}(j-\mu_2)^2 \sum_{j=0}^{N-1} p_{ij} \quad (8)$$

✓ ***Energy:*** Offers the sum of the squared elements in the GLCM. Also called uniformity or angular second moment. It measures the textural uniformity that is pixel pair repetitions and detects the disorders in the textures. The energy reaches a maximum value corresponding to one.

$$Eg = \sqrt{ASM} \ , \qquad \text{where} \quad ASM = \sum_{ij} p_{ij}^2 \quad (9)$$

✓ ***Homogeneity:*** Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. Also called Inverse Difference Moment. It measures the homogeneity of the image because it takes higher values for smaller gray tone differences in pair elements. Homogeneity is more sensitive to the clear presence of almost diagonal elements in the GLCM. It has a maximum value when all elements of the image are the same. The homogeneity decreases if the contrast increases while the energy is kept constant [29] [31].

$$H = \sum_{ij} \frac{p_{ij}}{1+|i-j|^2} \quad (10)$$

✍ **Improved Local Binary Pattern (LBP_{riu2}):** An LBP encodes local grayscale difference information in textured images, which is detected by sampling grayscale difference values at a central point $\boldsymbol{c}$ and $\boldsymbol{p}$ points spaced equidistantly around a circle of radius $\boldsymbol{R}$. The LBP descriptor is computed as:

$$LBP_{P,R} = \sum_{P=0}^{P-1} s(g_P - g_c)2^P \quad , \qquad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x \leq 0 \end{cases} \tag{11}$$

Where g_c denotes the gray value of the central pixel, while g_P represents the gray value of the neighboring pixel. To make the LBP descriptor robust to rotation, a uniform and invariant rotation descriptor $LBP_{P,R}^{riu2}$ has been proposed, which can be formally defined as

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{P=0}^{P-1} s(g_P - g_c) & \text{if} \quad U(LBP_{P,R}) \leq 2 \\ P + 1 & \text{otherwise} \end{cases} \tag{12}$$

Where

$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{P=1}^{P-1} |s(g_P - g_c) - s(g_{P-1} - g_c)| \tag{13}$$

where the exponent $riu2$ represents the rotation invariant "uniform" patterns with U ≤ 2. Therefore, the mapping of $LBP_{P,R}$ to $LBP_{P,R}^{riu2}$ has only P + 2 separate output values, resulting in a much shorter histogram representation for the entire image [32].

II.2.2 Convolutional Deep learning technique

The performance of pattern recognition systems is strongly influenced by the accuracy of their feature extraction algorithm. Many works have demonstrated the use of deep learning techniques in pattern recognition, but there is little evidence of their use as feature extractors.

✍ **Deep learning purpose:** pattern recognition system systems typically incorporate classifiers with learning behavior, such as Support Vector Machine (SVM), Random Forest Transform (RFT), and Radial Basis Function (RBF). These types of system may encounter significant problems when the training feature vectors have a high correlation rate, which significantly affects the system accuracy during the classification phase. In order to remedy this drawback, the system must itself extract the feature vectors (an own extraction step), then the obtained feature vectors serve as learning vectors (training base) for the classifier used. The overall system, therefore, receives raw images as input instead of pre-extracted feature vectors. Also, multi-level image analysis can provide accurate feature vectors because it is to behave as a multimodal method so that a feature-level fusion concept is applied when passing from an analysis level to another. Of course, the need to decorrelate the different feature

vectors of different classes (inter-class) is more important to obtain efficient pattern recognition systems, which requires the use of such deep methods.

✍ **Deep convolution learning principle:** Deep learning has many advantages over traditional machine learning techniques or other methods outside of machine learning. Although many problems can be solved by (learned) hand-craft methods, there are cases where these methods cannot cover the complexity of the task enough to be truly effective. The biggest advantage of deep learning is its flexibility, which is usually obtained from the number of layers of analysis that can be increased to achieve amazing complexity. We can use these techniques to analyze the image, under several levels, then extract the discriminating image feature.

The main difference between hand-crafted technique and deep learning lies in the analysis part. In fact, these two methods agree in the classification part, which is usually an intelligent classifier. In machine learning, the image's features are extracted independently of the classifier, which can affect its result if the feature extraction method used is inappropriate (gives feature vectors with a high inter-class correlation). On the contrary, in deep learning, the extracted feature vectors are automatically adapted to the classifier. Besides, in deep learning, the image is analyzed on several levels, which allows penetration into the image depth and extracting the appropriate features for its classification.

As shown in (Fig. II. 4), the analysis part consists of three secondary sub-parts, which are *i)* the convolution sub-part, in which the image is analyzed using several filters in order to capture only the important features, *ii)* the pooling sub-part, whose main objective is to reduce the volume of data by eliminating less important data. Finally, *iii)* the flatten sub-part, which allows to normalize the final feature vector to be appropriate at the input of the fully connected layer.

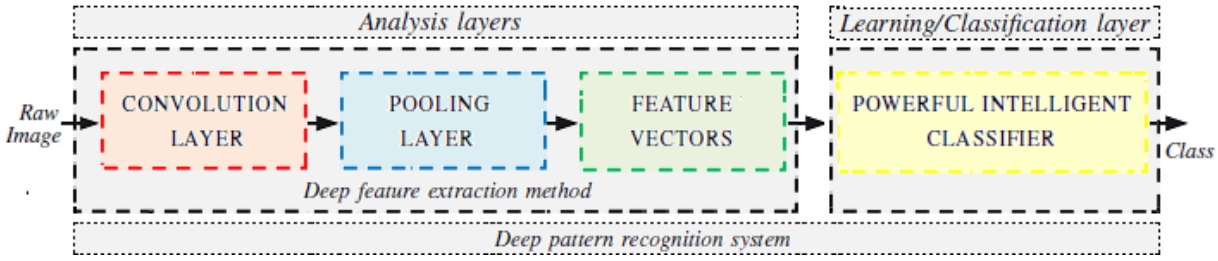


Fig. II.4. Image classification based deep learning technique

It should be noted that the image can be analyzed under several levels, so that each level can be composed of a convolution sub-part followed by a pooling sub-part. The length of the

final entity feature vector is mainly related to the pooling of the last level, which is related to the number of analysis levels and the number of convolutional filters in each level.

II.2.3 Feature selection technique

Feature selection is the process of finding features that are useful to represent the data and removing irrelevant features that contain redundant information. Feature selection methods aim to reduce the number of input variables to those considered most useful to a model in order to predict the target variable. Some predictive modeling problems have a large number of variables that can slow model development and training and require a large amount of system memory. In addition, the performance of some models may degrade when including input variables that are not relevant to the target variable. The feature selection methods can be classified into supervised and unsupervised methods. Unsupervised feature selection techniques ignore the target variable, such as methods that remove redundant variables using correlation. Supervised feature selection techniques use the target variable, such as methods that remove irrelevant variables [33].

There are many feature selection algorithms that can be used, such as Fisher Score, which is one of the most supervised and widely used feature selection methods. The main objective of the Fisher score (FS) is to obtain a set of data characterized by the divergence of the features of the different classes and the convergence of these features within the same class. In other words, this technique chooses the data that will increase the correlation of features of the same class and decorrelate them in different classes. Therefore, using the FS, we can calculate a score (Si) for each ith features

$$S_i = \frac{\sum_{j=1}^{C} n_j(\mu_{ij}-\mu_i)^2}{\sum_{j=1}^{C} n_j p_{ij}} \qquad (14)$$

Where C is the number of classes, μ_{ij} and ρ_{ij} are respectively the mean and the variance of the i^{th} feature in the j^{th} class, and n_j is the number of instances in the j^{th} class, and μ_i is the mean of the i^{th} feature. For the feature selection process, we calculate the Fisher scores (S_i) for all the features, then arrange these features in descending order based on the S_i values, then select the features with the highest values, while those with smaller S_i values are suppressed [34].

II.3 Proposed feature extraction method

Feature extraction from an image is the process of transforming the images raw pixel values into more meaningful and useful information that can be used in a machine vision

application. The rapid development of research has led to the emergence of new methods capable of deeply analyzing images and efficiently extracting dominant features, and the most important of these methods are those based on convolution. In this section, we consider the issue of CAD systems and introduce a new feature extraction method to efficiently represent the bone image feature. Thus, (Fig.II.5) shows the block diagram of the deep features extraction method from a bone image for use in the osteoporosis detection system.

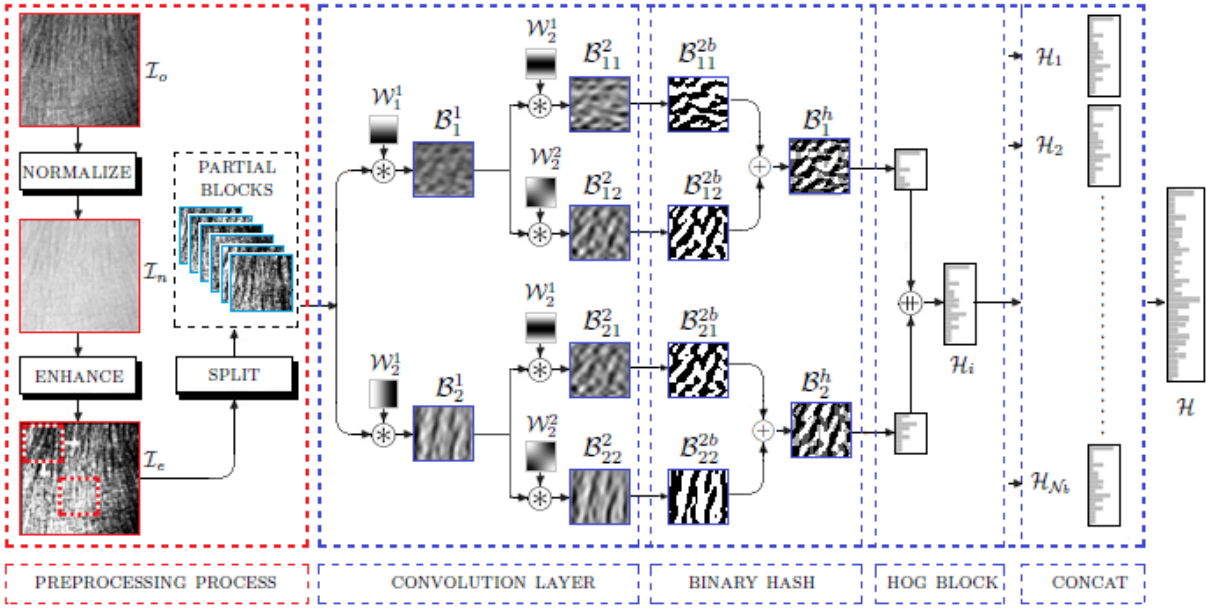


Fig. II. 5. Proposed bone feature extraction method using OS-PCANet deep learning. An example of a 2-stages OS-PCANet structure with 2 convolution filters in each stage.

The following sub-parts detail the basic concepts of our feature extraction method. [34]

II.3.1 Preprocessing Step

As the osteoporosis images obtained (raw data) cannot be used directly by detection systems, they must undergo a pre-processing during which a Region of Interest (ROI) is extracted. The role of this phase is to configure and modify the original image in order to prepare it for the feature extraction step. The ROI was localized by two anatomical landmarks (see annex B), then a square region, of fixed $N \times N = 400 \times 400$-sized pixels, was extracted. The original image is encoded in 16 *bits*, for this in our work, a normalization process is applied to transform the image's pixel into [0··255]:

$$I_n(i,l) = \frac{255}{\rho} \cdot I_o(i,j) \in [0 \cdot\cdot 255] \qquad (15)$$

where I_n and I_o denote respectively the normalized and original images, ρ represents the max value of the original image and $i,j = 1,2,\cdots,N$. Additionally, before starting the

feature extraction process, histogram equalization for the normalized image is applied to distribute the intensity values over the entire range.

$$I_n = \mathcal{F}_{CLAHE}(I_n) \quad (16)$$

where Ie denotes the equalized image and $\mathcal{F}_{CLAHE}(\circ)$ the equalization function. Thus, $\mathcal{F}_{CLAHE}$ is a method that supports low contrast, it operates on small regions of the image, called tiles, rather than the entire image. Neighboring tiles are then combined using bilinear interpolation to remove the artificial boundaries. This algorithm can be applied to improve the contrast of images.

In our method (see Fig. II.5), the image is analyzed block-by-block, and the feature vector of each $b \times b$-sized block is extracted by the OS-PCANet technique. Thus, in the splitting layer, the input $N \times N$-sized image is divided into square and overlapped blocks (B), each of which is then mapped into a feature vector via the OS-PCANet. The number of analyzing blocks (n_v) extracted from each image is equal to:

$$n_v = n \times n = \left\lfloor \frac{N-o}{b-o} \right\rfloor \cdot \left\lfloor \frac{N-o}{b-o} \right\rfloor \quad (17)$$

where o denotes the horizontal/vertical overlap between two adjacent blocks and $\lfloor \alpha \rfloor$ is the integer part of the value α. Finally, the entire image (I_e) is represented by a set of blocks ($\mathcal{S}_b$) as follows:

$$\mathcal{S}_b = \{B_{ij}\}_{i,j=1}^{n} \quad (18)$$

In our feature extraction method, each block is analyzed individually, then all the feature vectors extracted from all the blocks are concatenated into a single vector in order to represent the overall feature of the entire image.

II.3.2 Feature Extraction Step

In this step (see Fig. II.5), the feature vector of the image is extracted by extracting all the feature vectors of the blocks. Prior to this process, the system first forms the convolution filters ($\mathcal{W}_i^{\ell}$) using the PCA technique.

✍ **Convolutional Filters Formation:** Our system uses a bank of convolutional filters whose composition differs from one stage to another, and this difference resides mainly in the training database. In fact, the method of forming these filters is the same for all stages, in which it is based on the PCA technique. In each stage (ℓ), the training images (B_i) are used to create L_ℓ , $\ell = 1,2,..,L$ convolutional filters (L is the number of stages). So, for each image

and for each pixel of this image, we use the patch of size $k^{\ell} \times k^{\ell}$ (square filters) to truncate the neighbors of this pixel into an array $k^{\ell} \times k^{\ell}$, then reshape them into a vector, ϑ, of size $k^{\ell} \cdot k^{\ell} \times 1$. All the column vectors extracted from the image B_i are concatenated to obtain a 2D-vector $\mathcal{V}_i$:

$$\mathcal{V}_i^{\ell} = \{\vartheta_j\}_{j=1}^{\ell}, \qquad i = 1,2,..,\mathcal{T}_r \tag{19}$$

where $\mathcal{T}_r$ denotes the number of training images,ϑ_j ($j = 1,2,..,\xi$) represents the j^{th} vectorized patch in image B_i and ξ is the number of extracted vectors that relate to the path size and to the overlap between the truncated patches. This process is applied to all the images of the training database to obtain the training vector for the PCA technique:

$$\mathcal{V}^{\ell} = \{\mathcal{V}_i^{\ell}\}_{i=1}^{\mathcal{T}_r} \tag{20}$$

After performing the PCA technique on $\mathcal{V}^{\ell}$, one obtains the L_{ℓ} principal eigenvectors:

$$\mathcal{U}^{\ell} = \{u_i^{\ell}\}_{i=1}^{L_{\ell}} \tag{21}$$

Finally, the PCA filters of stage ℓ can be obtained by reshaping the principal eigenvector u_i^{ℓ} ($i = 1,2,..,\ell$) as follows:

$$\mathcal{W}_i^{\ell} = \mathcal{F}_{k_{\ell} \times k_{\ell}}(u_i^{\ell}), \;\; i = 1,2,..,L_{\ell} \tag{22}$$

where $\mathcal{F}_{k_{\ell} \times k_{\ell}}$ is a function that maps the 1D-vector u_i^{ℓ} to a 2D-vector $\mathcal{W}_i^{\ell}$. It should be noted that the outputs of stage ℓ ($\ell \in [2 \cdots L]$) represent the training images of stage $\ell + 1$. As for the first stage ($\ell = 1$), its training images are the original images (raw images) [34].

✍ **OS-PCANet Feature Extraction:** After having configured the convolutional filters, it remains to extract the features of the input image. Fig. II.5 shows an example of a two-stage system, in which it is obvious that the feature extraction process consists of three successive layers: convolution, binary hashing, and histogramming.

- ***Convolution layer:*** This layer consists of several stages (in our example $L = 2$). The outputs of each stage are obtained by filtering the input images by the convolutional filters previously constructed. The outputs of the first stage ($\ell = 1$) is obtained as:

$$B_i^1 = B * \mathcal{W}_i^1, \;\; i = 1,2,..,L_1 \tag{23}$$

where B_i^1 represents the filtered output images of first stage and the symbol $*$ denotes a 2D convolution process. The images which were filtered in the first stage are also subjected to a filtering process in the second stage as follows.

$$B_{ij}^2 = B_i^1 * \mathcal{W}_j^2, \quad i = 1,2,..,L_1, \quad j = 1,2,..,L_2 \tag{24}$$

Finally, for each input image, we can obtain $L_1 \cdot L_2$ filtered images at the output of the convolution layer.

- ***Binary hashing layer:*** This layer contains two main steps. The first is to convert all the outputs of the convolutional layer (last stage) to binary images:

$$B_{ij}^{2b} = \mathcal{F}_{HVSD}\left(B_{ij}^2\right), \quad i = 1,2,..,L_1, \quad j = 1,2,..,L_2 \tag{25}$$

Where $\mathcal{F}_{HVSD}$ is the Heaviside step function which returns the value "1" for positive inputs, and "0" otherwise. The role of the second step is to reduce the amount of data, in which the L_2-bits of each position are converted to an integer value:

$$B_i^h(x,y) = \sum_{j=0}^{L_2-1} B_{ij}^{2b}(x,y) \cdot 2^j, \quad i = 1,2,..,L_1 \tag{26}$$

After decoding each L_2 group separately, we get L_1 integer-valued images. It is important to note that whatever the number of stages, the number of outputs of this layer is equal to the number of filters used in the first convolution stage.

- ***Histogramming layer:*** The role of this layer is to extract the feature vector of the input image, for this the histogram of each hashed image (among L_1 images) is calculated and then concatenated to form a 1D feature vector. Thus, each hashed image (B_i^h) is partitioned into $\mathcal{N}_b$ overlapping blocks and a HOG histogram in each block (B_s) is computed.

$$h_s = \mathcal{F}_{HOG}(B_s), \quad s = 1,2,..,\mathcal{N}_b \tag{27}$$

where $\mathcal{F}_{HOG}$ is a function that performs the HOG technique. The length of this histogram (δ) depends on the size of the block (B_s) and the overlap rate used. Thus, the feature vector of each hashed image is obtained by concatenating all B_s histograms such as:

$$\mathcal{H}_i = [h_1, h_2, h_3, .., h_{\mathcal{N}_b}], \quad i = 1,2,..,L_1 \tag{28}$$

Finally, after applying this process in all the hashed images, the feature vector of the input image B is then defined as:

$$\mathcal{H} = [\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, .., \mathcal{H}_{L_1 \cdot \mathcal{N}_b}] \tag{29}$$

It is important to note that the length and precision of the vector ($\mathcal{H}$) change depending on the HOG parameters [34].

II.4 Conclusion

Deep learning is an emerging technology that has nowadays reached a substantial level of development, enabling them to have already a practical impact on several research fields. Computer vision and image processing are especially gaining momentum, due to the latest innovations in deep learning and improved accessibility of computational resources, such as powerful GPUs. This is why, to follow this evolution, we have proposed in this chapter a CAD system for osteoporosis based on Lightweight and efficient deep technique.

Chapter 3

Experimental Results and Discussion

Analyzes and Discussions

Abstract

In order to effectively represent the features of the bone image, the captured images are usually analyzed by the feature extraction method. The new representation produced by this feature extraction method must be unique to each case (patient/healthy). In fact, this chapter aims to clarify the feasibility and effectiveness of our proposed scheme through a detailed description of its performance, and hence the performance of the proposed automatic osteoporosis detection scheme.

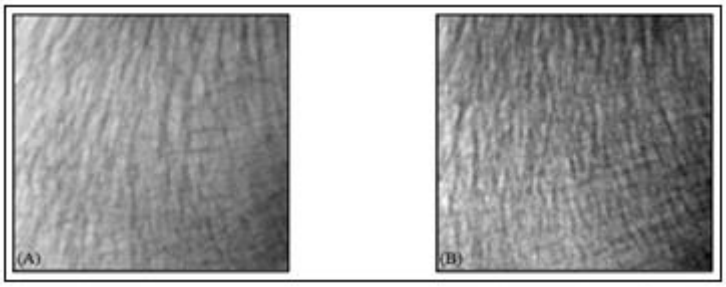
Experimental Results and Discussion

Analyzes and Discussions

Currently, a number of clinical decision tools have been developed for osteoporosis risk assessment by measuring bone mineral density and/or analyzing bone images. Fortunately, deep learning has proven its great ability to represent images and therefore to distinguish them. In this chapter, to assess the effectiveness of the proposed automated system for the detection of osteoporosis, we will examine its performance using an approved database.

III.1 Dataset description

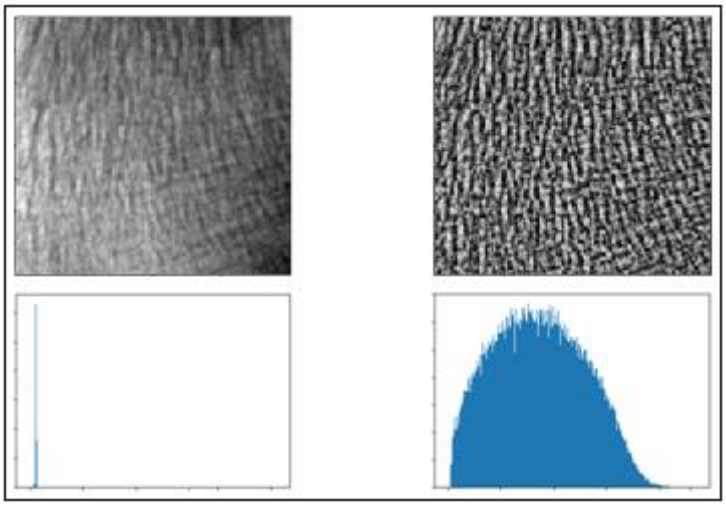
Our experiments were conducted on an available and publicly osteoporosis dataset provided during the IEEE-ISBI Challenge: Bone Texture Characterization [35] which contains images of osteoporosis acquired using a dual X-ray absorptiometry. This database consists of a set of 174 textured images, half of which from osteoporotic patients (with fracture) and the other half from control subjects. The fractures are distributed as follows: 22 neck fractures, 22 vertebral compaction, 22 wrist fractures, and 21 other fractures.



(a) *(b)*

Fig. III.1. Two different samples of the database used. *(a)* Healthy sample, and *(b)* Sample with osteoporosis

All the images in the database were subjected to normalization according to the CLAHE method. This method contains several parameters, the most important of which is the size of the analysis block, for this, empirical tests were carried out to determine this size which is finally fixed at 40×40. The following figure shows the results of the applied process.



(a) *(b)*

Fig. III.2. Bone image normalization. *(a)* Original image, and *(b)* Normalized image

This figure (Fig. III.2.) shows that the bone image is improved after the normalization process, as the dominant features to be extracted are clearly visible, which plays a major role in the classification process.

III.2 Test protocol

To evaluate the performance of the proposed CAD system, we have adopted two ways, the first is based on entire image analysis, while the second uses block-based image analysis. Additionally, we have applied the principle of feature selection in both ways. Furthermore, a 5-fold cross-validation procedure was used, where we divided the entire database into 5 folds each containing 34 samples, half (17) of which are healthy subjects, and the other half are osteoporotic subjects.

Cross-validation is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model and the other used to validate the model. In typical cross-validation, the training and validation sets must cross-over in successive rounds so that each data point has a chance to be validated against. The basic form of cross-validation is k-fold cross-validation. Other forms of cross-validation are special cases of k-fold cross-validation or involve repeated cycles of k-fold cross-validation.

In k-time cross-validation, data is first divided into k segments or folds of equal (or nearly equal) size. Subsequently, k iterations of training and validation are performed so that at each iteration a different fold of the data is kept for validation, while the remaining $k - 1$ folds are used for training. (Fig. III.3) shows an example with $k = 3$. The darker section of the data is used for training while the lighter sections are used for validation [36].

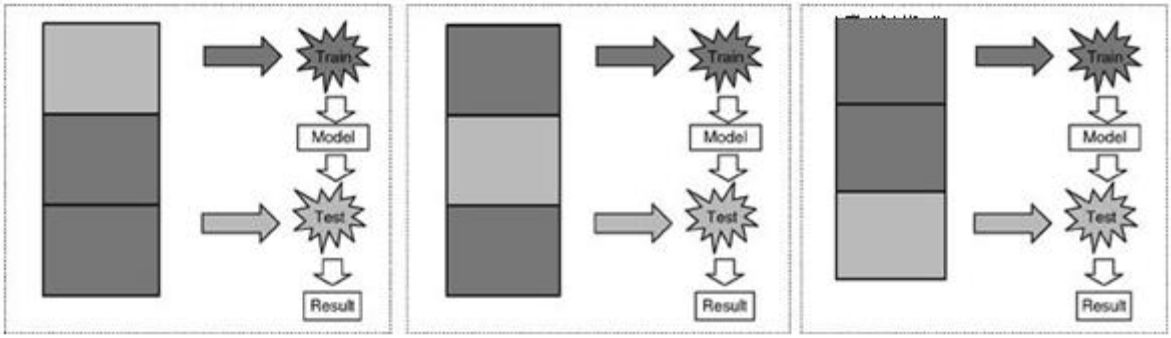


Fig. III.3. Procedure of three-fold cross-validation

III.3 Performance Evaluation

For each machine learning or deep learning model, we need to know how much the model has learned from the training data. Additionally, we need to know how well the same model predicts future or unseen data.

III.3.1 Metrics

For evaluation, we need a way to measure the model performance; these measurements are called evaluation metrics, there are many of them while the most popular are:

✍ ***Accuracy:*** The accuracy metric (Acc) measures the ratio of correct predictions to the total number of instances evaluated.

$$\text{Acc} = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

Where TP and TN denote true positive and true negative, and FP and FN denote false positive and false negative. TP is an outcome where the model correctly predicts the positive class. Likewise, a TN is an outcome where the model correctly predicts the negative class. A FP is an outcome where the model incorrectly predicts the positive class. Finally, a FN is an outcome where the model incorrectly predicts the negative class.

✍ ***Precision:*** Precision is used to measure positive patterns that are correctly predicted from the total of predicted patterns in a positive class.

$$p = \frac{TP}{TP+FP} \quad (2)$$

✍ ***Recall:*** The recall is used to measure the fraction of positive patterns that are correctly classified.

$$r = \frac{TP}{TP+TN} \quad (3)$$

✍ ***F-Measure (F1 Score):*** This metric represents the harmonic mean between the recall and precision values.

$$FM = \frac{2*p*r}{p+r} \quad (4)$$

III.3.2 Confusion Matrix

In supervised machine learning, the confusion matrix is a matrix that measures the performance of a classification model, on a set of test data.

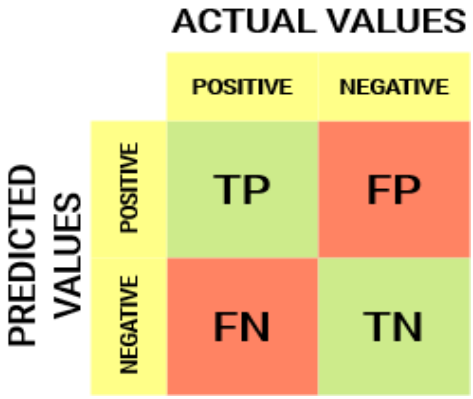


Fig. III.4. An example of a confusion matrix

III.3.3 Roc curve

A Receiver Operator Characteristic (ROC) curve is a graphical plot used to show the diagnostic ability of binary classifiers. It was first used in signal detection theory, but is now used in many other fields such as medicine, radiology, natural hazards and machine learning.

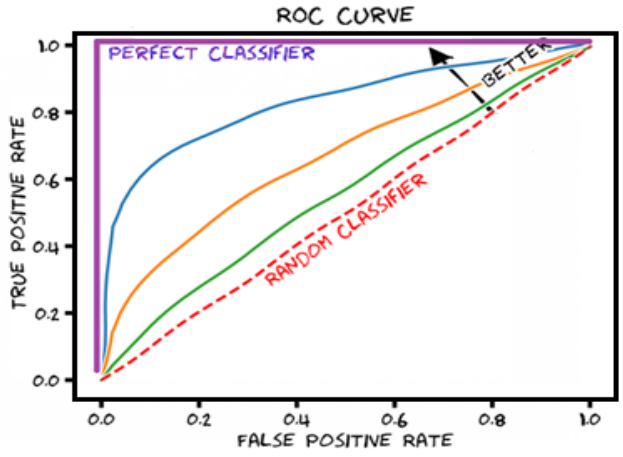


Fig. III.5. Roc curve for test data set

A ROC curve is constructed by plotting the true positive rate (TPR) against the false positive rate (FPR). The TPR is the proportion of observations that were correctly predicted to be positive out of all positive observations. Similarly, the FPR is the proportion of observations that are incorrectly predicted to be positive out of all negative observations [37].

III.4 Performance Tests Results

In this study, the test set was divided into two main parts according to the type of feature extraction (handcrafted or deep). The first part is about the performance of the system based on the handcrafted features, while in the second part we assess the performance of the system based on the deep features.

III.4.1 Hand-crafted feature

Before starting to test the proposed osteoporosis detection system (effectiveness of deep learning compared to other methods), a brief description of how to extract features using Hand-crafted feature extraction methods will first be given.

For LBP_{riu2}, (see Fig. III.6) we built three different histograms from each frame/block which have their own parameters, and then we combined them into one to get the feature vector.

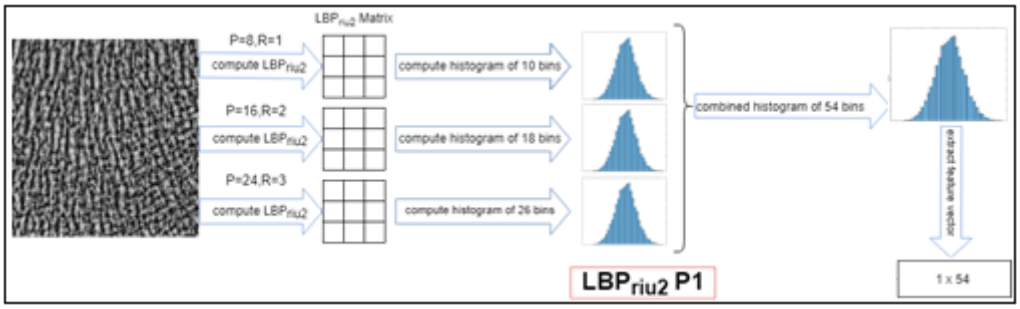


Fig. III.6. LBP_{riu2} based feature extraction method

For the GLCM features, we first calculated 9 GLCMs for each image/block, all of them are symmetrical with a size of 256×256 each of them has its own combination of distance and angle parameters, then we extracted four features of each GLCM which are contrast, correlation, homogeneity and energy. As a result, for each image/block, we obtained a feature vector with 36 features (see Fig. III.7).

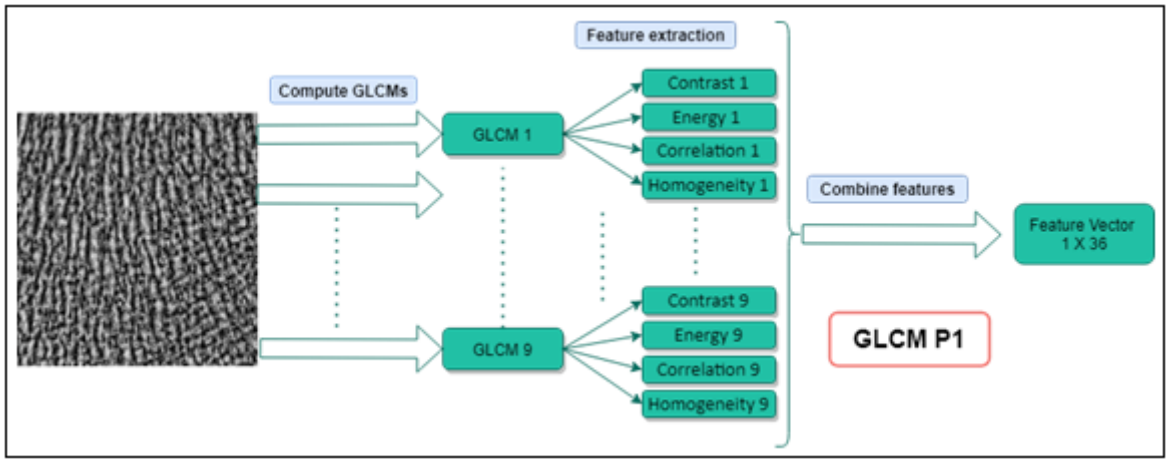


Fig. III.7. GLCMs based feature extraction method

For SIFT and ORB, we first extracted the descriptors of N keypoints obtained from each image, then we fed them all to k-means which gave us the codebook with k number of clusters, and finally we created an image histogram for each based on the codebook (we used multiple N and k values). As a result, we obtained feature vectors of 300 features with $N = 25$ and $k = 300$.

In this part, we will evaluate the Hand-crafted based feature extraction methods used when we use the whole image or analyze it in blocks. In both cases, we used the SVM classifier.

✍ **Whole image analysis:** Using the four methods (LBP_{riu2}, ORB, SIFT and GLCM) with different parameters, we can get the feature vector of the input image. This vector can be used directly or after selection of the dominant feature, which can be achieved by the Fisher method. Thus, a comparison between the different feature extraction methods used is presented in (Table III.1.) This table shows the performance of the osteoporosis detection system based on accuracy, respecting the parameters of each feature extraction method.

Table III.1. Osteoporosis detection system performance using whole Image analysis

Extraction methods	**Parameters**		**Number of features**	**Accuracy**	
				Without FS	With FS
LBP_{rui2}	R=1, P=8		54	66%	70% (20%)
	R=2, P=16				
	R=3, P=24				
GLCM features	D=1	Theta=0	36	64%	66 % (70%)
		Theta=π/2			
		Theta= π /4			

	D=3	Theta=0			
		Theta= π /2			
		Theta= π /4			
	D=5	Theta=0			
		Theta= π /2			
		Theta= π /4			
BoVW (Sift+Kmeans)	K=100	Keypoints=50	100	51%	/
		Keypoints=100		59%	68%
	K=300	Keypoints=50	300	60%	/
		Keypoints=100		54%	/
BoVW (ORB+K-means)	K=100	Keypoints=50	100	54%	/
		Keypoints=100		57%	68 %
	K=300	Keypoints=50	300	53%	/
		Keypoints=100		49%	/

From (Table III.1.), we can observe a good performance when using the whole feature vector (without feature selection) of the feature extraction method based on LBP_{riu2} parameterized by R = [1, 2, 3] and P = [8, 16, 24]. In this case, the osteoporosis detection system can operate with a performance up to ACC = 70% and ACC = 66%.

In the feature extraction method based on GLCM, we get the best result (ACC = 66% and ACC = 64% with and without feature selection, respectively) using the parameters D = [1,3,5] and theta = [0, π /2, π /4]. On the other hand, the BOVW method (SIFT-ORB), for $k = [100,300]$ and keypoints = [50,100], gives an acceptable ACC of 68% and 60% with and without fisher, respectively. It is important to note that to select the dominant feature of SIFT and ORB, we used 100 keypoints and K = 100 because these parameters gave us the highest average ACC for models without feature selection.

The feature extraction method based on LBPrui2 had the best performance for this type of analysis (whole image analysis). It can achieve ACC of 66% and 70% without and with feature selection task, respectively. Finally, from these results, we can see the poor performance of the feature extraction method based on ORB.

✍ **Block-based analysis:** In the next part of our study, we will try to improve the performance of the system using the block analysis method. Therefore, the image will be divided into blocks of equal size, and after extracting the features (using the best parameters

selected previously in the first part of the tests), we form the overall vector of the image using the concatenation of the different vectors of each block. Then, the performance of the system is evaluated in the same way as in the previous part. To limit the number of tests, we only divide each image into 4, 16 and 64 blocks. Thus, the results obtained for these decompositions are presented in (Table III.2).

Table III.2. Osteoporosis detection system performance using block-based analysis

Extraction methods	Number of blocks	Number of features	Accuracy	
			without FS	with FS
LBP	2x2	216	66%	67%
	4x4	864	64%	68%
	8x8	3456	64%	75%
GLCM	2x2	144	65%	66%
	4x4	576	64%	70%
	8x8	2304	68%	75%
BoVW (SIFT + k-means)	2x2	300	56%	74%
	4x4	300	54%	80%
	8x8	300	49%	82%
BoVW (ORB + k-means)	2x2	300	59%	79%
	4x4	300	49%	80%
	8x8	Didn't detect enough keypoints in the block		

From the previous table (Table III.2.), GLCM features provided the best performance without feature selection, with an ACC of 68%. Whereas with the feature selection, SIFT had an ACC of 82%, which was unexpected given the poor performance of SIFT without feature selection. According to these experimental results, the osteoporosis detection system performance can be effectively improved as the number of analysis blocks increases.

III.4.2 Deep feature

To assess the effectiveness of the proposed osteoporosis detection system, in this subpart we will examine its performance using the database previously used with the same testing protocol. Our system includes several parameters to control the accuracy of the extracted feature vector, the most important of which are the number of stages in the convolution layer, the number and size of convolution filters per stage. It is therefore essential

to evaluate the performance of the system to select the appropriate parameters that provide the best results.

In general, in these cases, empirical tests are performed by modifying the various parameters in a predefined set and then selecting the combination that improves the classification rate. Before beginning, it should be noted that we have configured our system into one stage ($L = 1$). Further, the parameters used to calculate HOG histograms, which are the number of HOG windows, the number of histogram bins, the number of regions in the image, and the regions overlap ratio are fixed at 13, 9, 40 and 50%, respectively.

✍ **Whole image analysis:** In the preliminary tests (experimental setup), we tried to choose the number of filters (L_1) and the size of the filters ($k_1 \times k_1$), among the set of values 2, 4, 6, 8, and 11×11, 13×13, 15×15, 17×17, respectively. Since varying these parameters can produce many feature representations, we can experimentally select a combination (L_1, $k_1 \times k_1$) that can effectively improve the accuracy of the system by changing one of them each time and choosing the best one that gives the best performance. Thus, in order to see the effect of these parameters on the performance of the system, we clearly illustrate, in (Table III.3.), the classification results (AUC and ACC) in the case where the image is fully analyzed.

Table III.3. Experimental results of PCANET parameter selections

Filters Sizes	**2 filters**		**4 filters**		**6 filters**		**8 filters**	
	AUC(%)	ACC(%)	AUC(%)	ACC(%)	AUC(%)	ACC(%)	AUC(%)	ACC(%)
11 x 11	49	49	46	48	48	51	49	52
13 x 13	47	49	56	56	48	51	45	48
15 x 15	52	54	53	55	54	54	57	55
17 x 17	46	49	46	52	46	50	44	48

From this table we can extract two important remarks:

i) Regardless of the size and number of filters, the performance of the system is poor with all possible combinations due to the lower accuracy achieved (ACC) which is less than 56% in the best case.

ii) In general, medium sizes and a lower number of filters give high accuracy. Thus, the average ACC obtained by the combination (4, 13×13) is better than that obtained by the combination (8, 17×17).

From (Table III.3), it is evident that the combination (L_1, $k_1 \times k_1$) equal to (4, 13×13) gives better results in terms of ACC (ACC = 56% with AUC = 56%).

Thus, (Fig. III.8) shows the Receiver Operating Characteristic (ROC) curve obtained in this case where the system could reach respectively 54.12% and 57.65% for the sensitivity and the specificity with a False Negative Rate (FNR) and a False Positive Rate (FPR) equal respectively to 45, 88% and 42.35%.

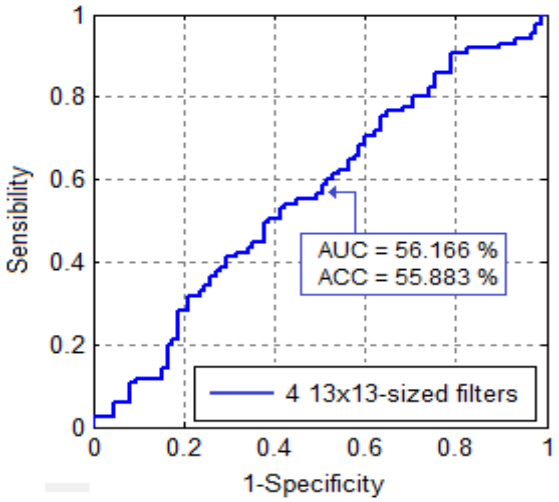
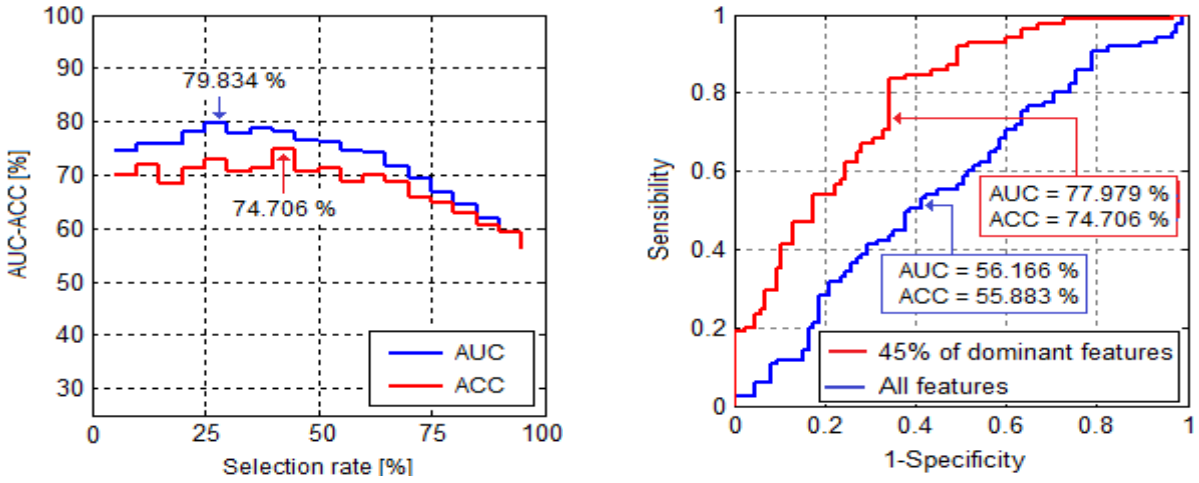


Fig. III.8. System performance under 4 filters of size 13×13 (all features have been used)

The major deterioration in system performance is due to the strong correlation between different vectors of different classes. Therefore, we will select the dominant components of feature vectors by applying Fisher's criterion. In this part, after sorting the components of the vectors, we will re-evaluate the performance of the system based on some components selected from the original vectors. In fact, we will choose from 10% to 100% with a step of 5% of the components of each vector then calculate the accuracy of the system. (Fig. III.9. *(a))* shows the performance of the system under all selection ratio. From this figure, the best result (ACC = 74.7059% and AUC = 77.9792%) is obtained using 45% of vector components, in which an improvement of 33.70% and 38.80% was obtained. Finally, (Fig. III.9. *(b))* clearly shows the positive impact of selecting features compared to using all features.



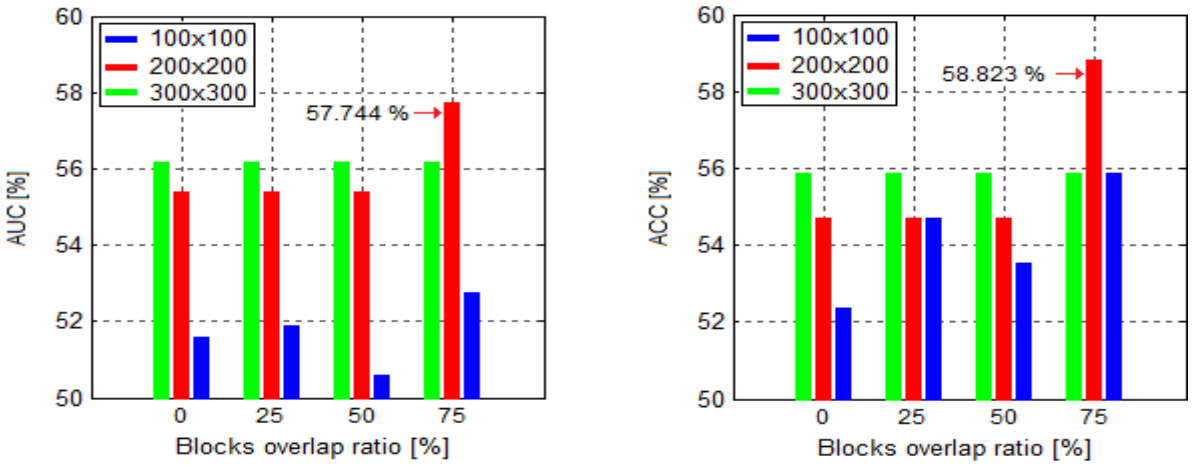
(a) *(b)*

Fig. III.9. Impact of the selection of dominant features on system performance. *(a)* System performance variation based on feature selection ratio, and *(b)* Performance comparison when using features selection

✍ **Block-based analysis:** Many results in the literature have proven the effectiveness of image/video processing using the division of the image into regions (blocks), for this we will present in this part the results obtained during the block-based osteoporosis images analysis. This method divides the image into several blocks (with or without overlap), then extracts the feature vector of each block, and finally, these vectors are combined (concatenated) to obtain the feature vector of the entire image. In this part, we will keep the best combination obtained (4 of 13×13-sized convolutional filters) and we will choose both the size of the block and the overlapping ratio, respectively, among the following predefined values: {100×100, 200×200, 300×300} and {0%, 25%, 50%, 75%}. (Table III.4.) and (Fig. III.10. *(a))* and (Fig. III.10. *(b))* illustrate the results obtained for AUC and ACC rates.

Table III.4. System performance results (block-based analysis)

Blocks Size	**0%-OVERLAP**		**25%-OVERLAP**		**50%-OVERLAP**		**75%-OVERLAP**	
	AUC(%)	ACC(%)	AUC(%)	ACC(%)	AUC(%)	ACC(%)	AUC(%)	ACC(%)
100×100	52	52	52	55	51	54	53	56
200×200	55	55	55	55	55	55	58	59
300×300	56	56	56	56	56	56	56	56



(a) *(b)*

Fig. III.10. Impact of block-based analysis on system performance. *(a)* Variation of the AUC and *(b)* Variation of the ACC.

A closer look at these results reveals that the performance of our system has improved overall, but it is not satisfactory. Obviously, using the 200×200-sized block with an overlap ratio of 75% resulted in a 2.81% and 5.26% system performance improvement in terms of ACC and AUC, respectively. In this case, the system works with 51.76%, 65.88%, 48.24%, and 34.12% for sensitivity, specificity, FNR, and FPR respectively. (Fig. III.11) shows the improvement in system performance when used a block-based image analysis.

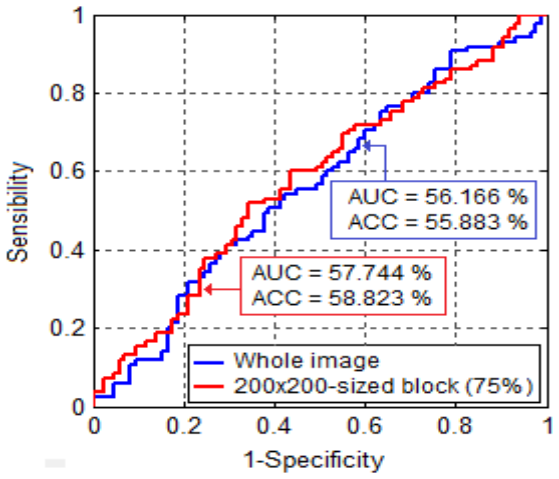
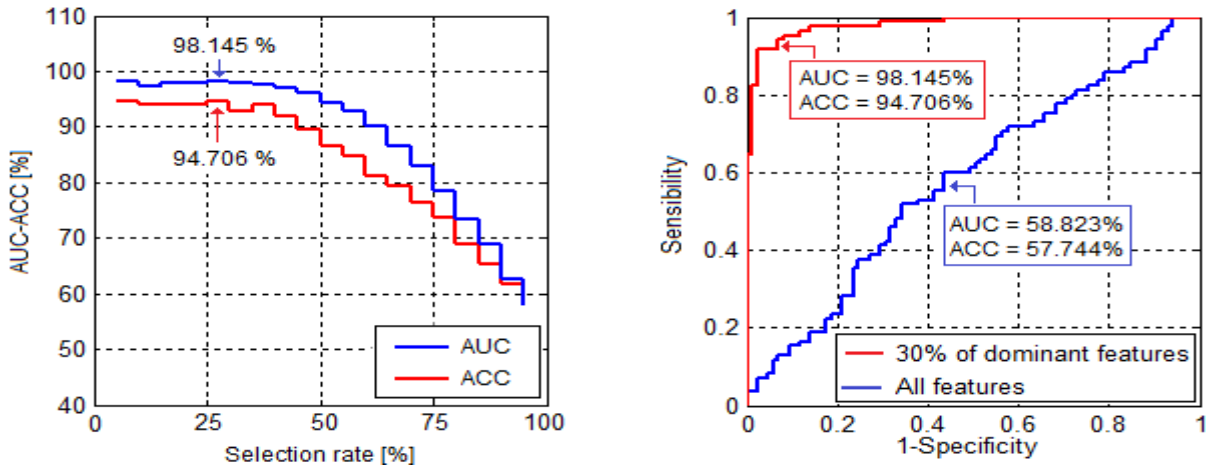


Fig. III.11. Performance comparison when using block-based analysis

To further prove our idea, the feature selection principle was applied in this case and the results obtained (for the 200×200-sized block) are shown in (Fig. III.12. *(a)*), from which it is very clear that first the performance of the system has been greatly improved and secondly, the best result has been obtained using 30% of vectors components. Indeed, compared to the

system performance without feature selection, improvements of 61.00% and 69.97% were obtained for the ACC and AUC, respectively, (see Fig. III.12. *(b)*).



(a) *(b)*

Fig. III.12. Impact of the selection of dominant features on system performance (using 200×200-sized block). *(a)* System performance variation based on feature selection ratio, and *(b)* Performance comparison when using features selection.

Indeed, a very efficient result was obtained, as the system operates at a sensitivity of 95.29% and a specificity of 92.94%, which explains the low values of FNR and FPR which are estimated at 4.71% and 7.0%, respectively

Finally, to demonstrate the importance of block-based analysis and the feature selection process, in (Fig. III.13), we compare the best results obtained. From this figure, the feasibility of block-based analysis is evident, but what is more striking is the significant impact of the feature selection process, which significantly improved the performance of the system, so that the improvement rate is reached above 69%.

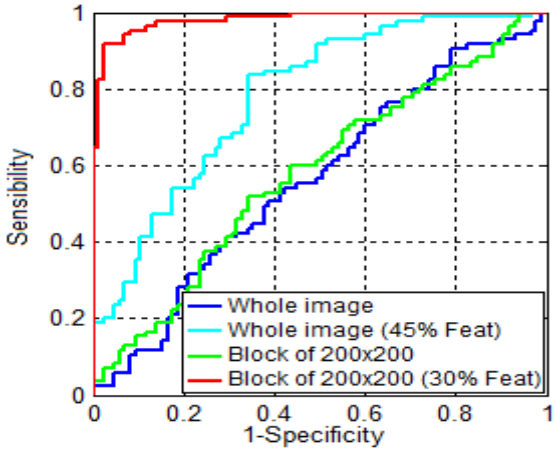


Fig. III.13. Performance comparison of all systems

It should be noted that it is also possible to improve this performance by using more than one stage in the convolution layer. Also, keep in mind that the HOG technique has several parameters (the number of HOG windows, the number of histogram bins, the number of regions in the image, and the regions overlap ratio) that can, with careful choice, improve the performance of the system for osteoporosis detection.

III.5 Conclusion

Deep learning is an emerging technology that has nowadays reached a substantial level of development, enabling them to have already a practical impact on several research fields. Computer vision and image processing are especially gaining momentum, due to the latest innovations in deep learning and improved accessibility of computational resources, such as powerful GPUs. This is why, to follow this evolution, we have proposed in this work a system for osteoporosis based on Lightweight and efficient deep technique. Thus, using an approved database and after a feature selection process, our proposed method has given excellent results compared with handcrafted methods. Logically, this excellent performance increases by increasing the number of stages in the convolution layer, and this is due to the possibility of accessing other features in the advanced stages of convolution.

General Conclusion

Conclusion

Unfortunately, due to osteoporosis, bone fracture is becoming an enormous health issue as millions of people around the world are affected by it. Thus, an early diagnosis of osteoporosis is important for the prevention of bone fractures because treatments are more effective in the early stages of the condition, before fractures have appeared. Currently, a number of clinical decision tools have been developed for osteoporosis risk assessment by measuring bone mineral density and/or analyzing bone images. Unfortunately, the great similarity and correlation between healthy and diseased bone images represents the major obstacle in building an effective osteoporosis detection system, making this task a major challenge. Unfortunately, the high similarity and correlation between healthy and diseased bone images represent a major obstacle to building an effective osteoporosis detection system, making this task a major challenge. But recently, new hope has emerged to establish an effective system, with the emergence of deep analysis.

The deep feature extraction method allows images to be analyzed in multiple levels to extract hidden features that can increase the accuracy of feature vectors, which in turn increases the pattern recognition system accuracy. In this dissertation, we will try to examine the effectiveness of a deep feature extraction method. To further illustrate our idea, a simple and less-complex method was chosen for use in low-performance devices (slow CPU and low memory), which is PCANet deep learning. Besides, an osteoporosis detection system, which is a prototype of a pattern recognition system, was used. Like all deep feature extraction methods, the method used analyzes the texture of the bone image at several levels. The experimental results demonstrate the effectiveness of the deep features compared to the features engineering (extracted with hand-craft methods).

This excellent performance increases by increasing the number of stages in the convolution layer, and this is due to the possibility of accessing other features in the advanced stages of

convolution. These results can be improved using the principle of data fusion, at the feature level, or at the score level, and this is what we see as future work using other lightweight deep learning techniques such as DCTNet, DSTNet, and ICANet.

Bibliographies

[1] Bones: Types, Structure and function. Medically reviewed by William Morrison, M.D. And written by Tim Newmanon, on Janury 11,2018

[2] Types of bones | Learn Skeleton Anatomy, V.Body. Medically reviewed by Lori Driscoll, MSLS, on July 25,2017

[3] Manolagas and Stavros C, Birth and death of bone cells: Basic regulatory mechanisms and implications for the pathogenesis and treatment of osteoporosis, Endocrine Reviews, Volume 21,115-37, 2000

[4] A.M. Parfit, Morphometry of bone resorption: Introduction and overview, Bone, Volume 14,435-441,1993

[5] Parfit, A.M., Problems in the application of in vitro systems to the study of human bone remodeling, Calcified Tissue International, Volume 56(1), 2015

[6] Brandi, M.L., Microarchitecture: The key to bone quality. Rheumatology, 48(Suppl-4), 2009

[7] NIH, National Institutes of Health, Consensus development panel on osteoporosis prevention, diagnosis and therapy. Jama, Volume 285(6), 2001

[8] Lorentzon, M., & Cummings, S. R. Osteoporosis: the evolution of a diagnosis. Journal of internal medicine, 277(6), 650-661, 2015

[9] Osteoporosis: Related Bone Diseases National Resource Center. The Surgeon General's report on bone health and osteoporosis NIH. Reviewed on June 26, 2017

[10] Osteoporosis. (n.d.) American Heritage® Dictionary of the English Language, Fifth Edition. (2011)

[11] Johnell O and Kanis JA. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures, Osteoporosis International, Volume 17(12),1726-33, 2006

[12] Kanis, John A, Diagnosis of osteoporosis and assessment of fracture risk. The Lancet, Elsevier, Volume 359(9321), 1929-1936, 2002

[13] Molino, Giulia, et al. "Imaging Techniques for the Assessment of the Bone Osteoporosis-Induced Variations with Particular Focus on Micro-CT Potential." Applied Sciences 10.24 (2020): 8939.

[14] Hume, P. A., & Ackland, T. Physical and clinical assessment of nutritional status. In Nutrition in the prevention and treatment of disease (pp. 71-84). Academic Press, 2017

[15] Blake, G. M., & Fogelman, I. Technical principles of dual energy x-ray absorptiometry. In Seminars in nuclear medicine (Vol. 27, No. 3, pp. 210-228). WB Saunders, 1997, July

[16] Papaioannou. A, Morin. S, Cheung. AM, Atkinson. S, Brown. JP, Feldmen. S, Hanley. DA, Hodsman. A, Jamal. SA, Josse. RG and others. Clinical practice guidelines for the diagnosis and management of osteoporosis in Canada: Background and technical report, Scientific advisory council of osteoporosis Canada, Ontario, 2010

[17] Berger and Abi, How does it work?: Bone mineral density scans. British medical journal, BMJ publishing group, Volume 325(7362), 482, 2002

[18] Kalpakcioglu, Banu. B and Morshed, Saam and Engelke, Klaus and Genant, Harry. K. Advanced imaging of bone macrostructure in bone fragility and fracture repair, JBJS, LWW, Volume 90 (S-1), 68-78, 2008

[19] Black, Dennis. M and Bilezikian, John. P and Ensrud, Kristine. E and Greenspan, Susan. L and Palermo, Lisa and Hue, Trisha and Lang, Thomas. F and McGowan, Joan. A and Rosen, Clifford. J. One year of alendronate after one year of parathyroid hormone for osteoporosis (1-84), New England Journal of Medicine, Mass Medical Soc, Volume 353(6), 555-565, 2005

[20] Sandra Avila. Extended Bag-Of-Words formalism for image classification. Computer Vision and Pattern Recognition [cs.CV]. Université Pierre et Marie Curie - Paris VI, 2013.English.

[21] Mohan, S. Image Classification using Bag of Visual Words Model. MLK - Machine Learning Knowledge, 2020, May 15

[22] Gunopulos D. Clustering Overview and Applications. In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA, 2009

[23] Singh, Himanshu. Practical Machine Learning and Image Processing. Apress, 2019.

[24] Lowe, D. G. Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91-110, 2004

[25] Data Enrichment. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA, 2017

[26] Rosten, E., Porter, R., & Drummond, T. Faster and better: A machine learning approach to corner detection. IEEE transactions on pattern analysis and machine intelligence, 32(1), 105-119, 2008

[27] Calonder, M., Lepetit, V., & Fua, P. Keypoint signatures for fast learning and recognition. In European conference on computer vision (pp. 58-71). Springer, Berlin, Heidelberg, 2008, October

[28] Kumar, D. Feature extraction and selection of kidney ultrasound images using GLCM and PCA. Procedia Computer Science, 167, 1722-1731, 2020.

[29] Patil, Maya Deoker1 Prof SN. "Osteoporosis Detection by using CT Images based on Gray Level Co-occurrence Matrix and Rule based approach." IJSRD - International Journal for Scientific Research & Development| Vol. 5, Issue 08, 2017 | ISSN (online): 2321-0613

[30] Hall-Beyer, M. GLCM texture: a tutorial v. 3.0 March 2017.

[31] CHIH-CHENG. HUNG. Image texture analysis: Foundations, Models and Algorithms. Springer, 2020

[32] Ojala, T., Pietikainen, M., & Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence, 24(7), 971-987, 2002

[33] Kuhn, M., & Johnson, K. Applied predictive modeling (Vol. 26). New York: Springer, 2013

[34] Meriem mebarkia, Abdallah Meraoumia, Lotfi Houam, Seddik Khemaissia, Bouchemha Amel and Rachid Jennane. Hierarchical Multiscale Local Binary Pattern for Better Osteoporosis Detection, Conference IEEE, Libya, 2021

[35] Challenge IEEE-ISBI: Bone Texture Characterization, available at: http://www.univorleans.fr/i3mto/challenge-ieee-isbi-bonetexturecharacterization

[36] Refaeilzadeh P., Tang L., Liu H. Cross-Validation. In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA, 2009

[37] Hanley, John A. Receiver operating characteristic (ROC) methodology: the state of the art. Crit Rev Diagn Imaging, Volume29(3), 307-335, 1989

[38] Nelli, F. Python data analytics: Data analysis and science using PANDAs, Matplotlib and the Python Programming Language. Apress, 2015

[39] Yu H. Support Vector Machine. In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA, 2009

[40] Patle, A., & Chouhan, D. S. (SVM kernel functions for classification. In 2013 International Conference on Advances in Technology and Engineering (ICATE), 1-9, IEEE, January, 2013

[41] Rüping, S. SVM kernels for time series analysis (No. 2001, 43). Technical report. 2001

[42] Pedregosa et al. Scikit-learn: Machine Learning in Python. JMLR 12, 2825-2830, 2011

[43] Lotfi, Houam & Hafiane, Adel & Jennane, Rachid & Abdelhani, Boukrouche & Lespessailles, Eric. (2010). Trabecular Bone Anisotropy Characterization Using 1D Local Binary Patterns. 105-113. 10.1007/978-3-642-17688-3_11.

[44] Lotfi, Houam & Hafiane, Adel & Abdelhani, Boukrouche & Lespessailles, Eric & Jennane, Rachid. (2012). One Dimensional Local Binary Pattern for Bone Texture Characterization. Formal Pattern Analysis & Applications. 17. 10.1007/s10044-012-0288-4

Support Vector Machines

Usually, a learning algorithm tries to learn the most common features (what differentiates one class from another) of a class and the classification is based on the representative features learned (therefore the classification is based on the differences between classes). There are many classification algorithms that belong to Machine Learning (ML) such as k-Nearest Neighbors, Random Forest, Stochastic Gradient Descent, Artificial Neural Network, and Support Vector Machine.

Each with its unique characteristics, specific to the nature of the data and the predictive model you want to build. The data to be subjected to a pattern during the training phase can be arrays composed of a single value per element, or of a multivariate value. These values are often called features or attributes. SVM is the most popular algorithm for classification.

A.1 Introduction

SVMs are a number of machine learning techniques that were first developed in AT&T laboratories by Vapnik and colleagues in the early 90s. The basis of this class of procedures is actually an algorithm called Support Vector, which is a generalization of a previous algorithm called Generalized Portrait, developed in Russia in 1963 by Vapnik [38]

SVM represents a set of supervised learning techniques that creates a model from training data, the learned model can be used to predict the output of a new object. During the learning phase, it projects the observations into a multidimensional space called decisional space and build a separation surface called the decision boundary which divides this space into two

belonging areas. SVMs are also used for regression, for which they are referred to as Support Vector Regression (SVR). There are two special properties of SVMs which are:

✍ ***High generalization:*** the performance of the learned function on test data or "unseen" (blind) data that is excluded from training by maximizing the margin (the distance between the hyperplane and the closest data vectors in feature space).

✍ ***Support efficient nonlinear classification*** by kernel trick which is a method to convert linear classifier to nonlinear classifier by using nonlinear function to map original observations in higher dimensional space [38] [39].

A.2 Kernel functions

The kernel function is a mathematical trick that allows the SVM to perform a two-dimensional classification of a set of originally one-dimensional data (feature vectors). In general, a kernel function projects data from a low dimensional space to a higher dimensional space. The selection of the kernel function depends on the application; it is not fixed [40].

✍ ***Linear Kernel Function***: The linear kernel $K(x, y) = (x).(y)$ is the simplest kernel function. The decision function takes the form $(x) = w.x + b$. When using the linear kernel to predict time series, it means that the resulting model is an autoregressive statistical model of order k.

✍ ***Radial Basis Function (nonlinear):*** Radial basis function (nonlinear): The radial basis kernels (Gaussians), take the form $\mathrm{K(x, y)} = \mathrm{e}^{-\gamma\|\mathrm{x}-\mathrm{y}\|^2}$ (kernel function parameter $\gamma > 0$). Clearly, the similarity of two examples is simply judged by their Euclidean distance.

✍ ***Sigmoid Function (nonlinear):*** Here $\mathrm{(x, y)} = \tanh(\gamma \mathrm{x}^{\mathrm{T}}\mathrm{y} + \mathrm{r})$. Sigmoid kernel or hyperbolic tangent kernel [41]. The (Fig. A.1) below shows an example of these kernels.

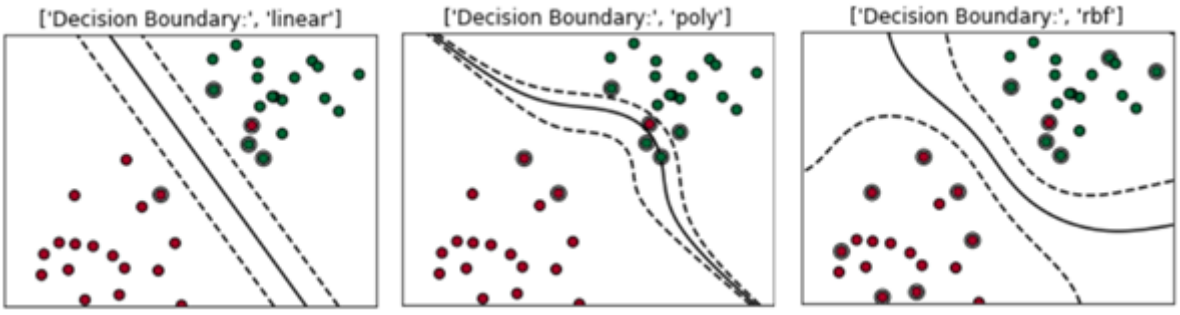


Fig. A.1. SVM with different kernel functions

✍ **Regularization (C parameter):** Adds a penalty for each misclassified data point. If it is small, the penalty for misclassified points is small, so that a decision limit with a large margin is chosen at the expense of a greater number of misclassifications.

The C parameter trades off the correct classification of training examples against the maximization of the decision function's margin (see Fig. A.2).

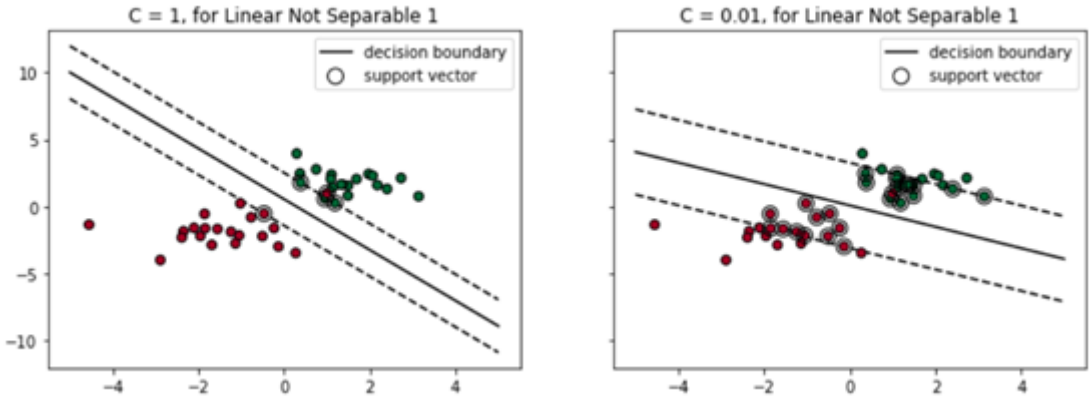


Fig. A.2. Example of two different C parameters C=1 and C=0.01

For larger values of C, a smaller margin will be accepted if the decision function is better at classifying all training points correctly. A lower C will favor a larger margin, therefore a simpler decision function, at the cost of training accuracy. In other words, C behaves like a regularization parameter in the SVM [42].

✍ **Gamma parameter:** The gamma parameter defines how far the influence of a single training example reaches. This means that high Gamma will consider points close to the plausible hyperplane and the low Gamma will consider points at a greater distance (see Fig A.3.). By decreasing the Gamma, finding the right hyperplane will take into account points at greater distances, so more and more points will be [42].

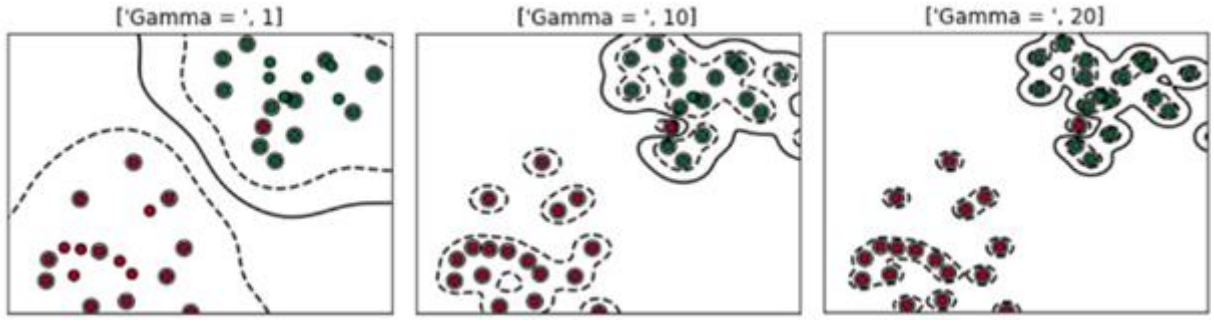


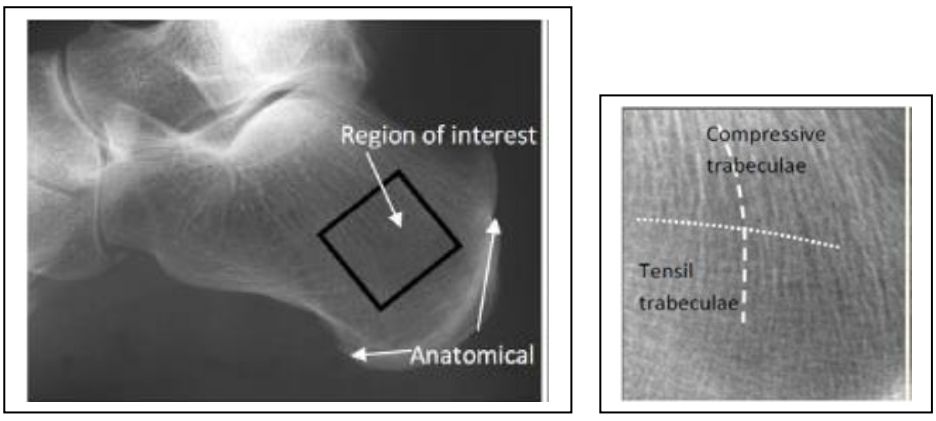
Fig. A.3. Example of three different Gamma parameters Gamma=1, Gamma=10 and Gamma=20

Region of Interest Extraction

For all subjects, radiographs of the calcaneus were obtained with a high-resolution X-ray device with direct digitization (BMATM, D3A Medical Systems, Orleans, France).

The calcaneus (heel bone) is well suited to measure the anisotropy (Fig. B.1 (a)). This bone is submitted to compression and tension forces produced by the walking and by the gravity. As a result, it is a very anisotropic structure as shown in (Fig. B.1 (b)). The evolution of the orientations of the trabeculae enables quantifying the degree of deterioration of the bone. For a normal subject both types of trabeculae are uniformly distributed.

For an osteoporotic subject the number of tensile trabeculae decrease gradually until a complete disappearance for a patient with severe osteoporosis. On the other hand, compression trabeculae become thinner and their number decreases much less quickly during the disease. As a result, a radiograph of an osteoporotic subject will be more anisotropic than the normal subject.



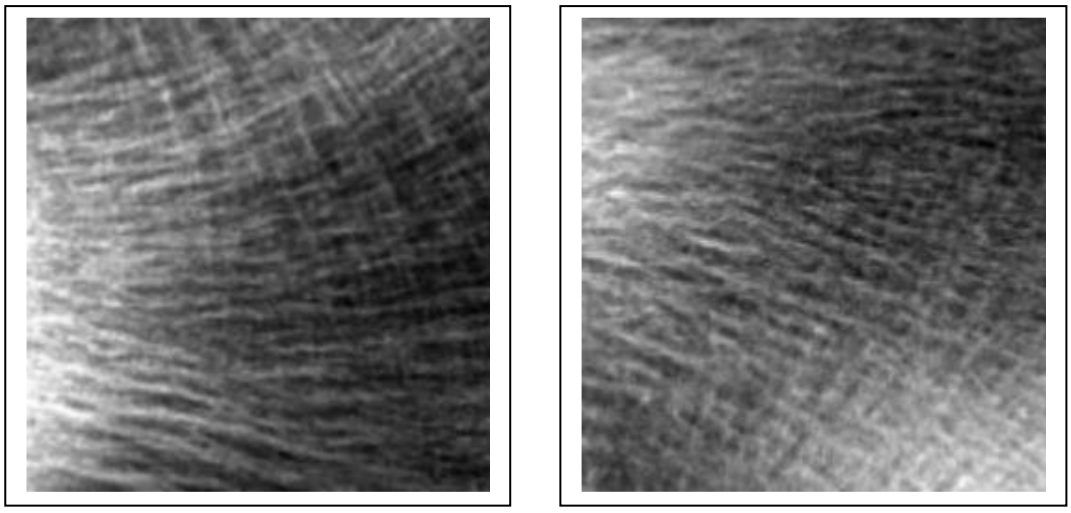
(a) (b)

Fig. B.1. Calcaneus radiograph and its region of interest (dark square). (a) Region of interest measuring 400×400 pixels used for processing and testing (b).

Calcaneus X-ray images were obtained following a standardized procedure, using X-ray clinical equipment with a tungsten tube and an aluminum filter of 1-mm thickness. Focal-calcaneus distance was set at 1 m. The calcaneus was placed in contact with the sensor; the tube voltage was fixed at 36 kV and the exposure condition was 18mA, with an exposure time of 80 ms.

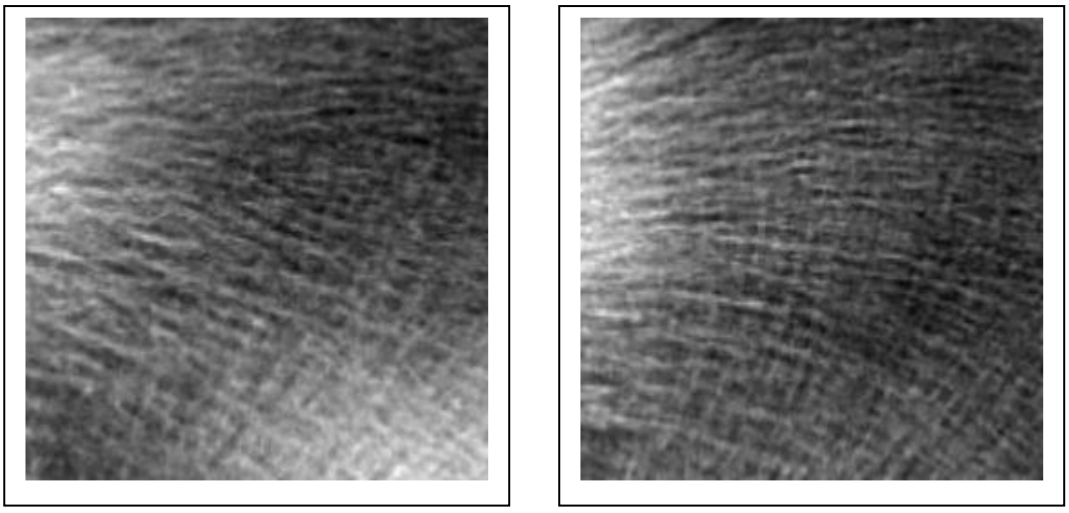
The region of interest (ROI) (see Fig. B.1 (a)) was defined by a physician by marking anatomical markers that can be easily identified on the calcaneus image. This ensures that the ROI (Fig. 1(b)) be acquired in the same area as well as in the same orientation from each bone radiography, since the effect of the orientation on the analysis is part of this study. This ROI of 400 x 400 was located in a part of the calcaneus that contains only trabecular bone. The pixel size was 105μm. [43] [44]

The following figures (Fig. B.2.) and (Fig. B.3.) presents four different radiographic image samples of the calcaneus. The first figure (Fig. B.2.) shows two images (Fig. B.2 (a)) and (Fig B.2(b)), which are from osteoporotic patients, while (Fig. B.3) shows two images (Fig. B.3(a)) and (Fig. B.3 (b)), which are from healthy subjects.



(a) (b)

Fig. B.2. Two radiographic images of the calcaneus, from osteoporotic patients. (a) and (b)



(a) (b)

Fig. B.3. Two radiographic images of the calcaneus, from healthy subjects. (a) and (b)

It should be noted that these calcaneus classes represent a challenging dataset, since all of them show similar visual characteristics.