Comparison of Algorithms to Categorize Breast Composition in Mammography Images

Júlia Sanàbria Franquesa

PROJECT REPORT

Supervisors

PROF. FILIPPO MOLINARI
ING. MASSIMO SALVI

Politecnico di Torino

Torino, August 22, 2020
Comparison of Algorithms to Categorize Breast Composition in Mammography Images

by

Júlia Sanàbria Franquesa

under the supervision of

Filippo Molinari, PhD
Massimo Salvi, PhD

This Master thesis has been conducted under the supervision of Prof. Filippo Molinari and Ing. Massimo Salvi mainly at Department of Electronics and Telecommunication (DET) at Politecnico di Torino; in collaboration with Dr. Christian Mata from Department of Chemical Engineering at Universitat Politècnica de Catalunya.

Torino, August 22, 2020
To my mum, Marta Franquesa i Niubó (in memoriam)
I would like to express my sincere gratitude to Dr. Massimo Salvi for giving me the opportunity to join the Politecnico di Torino under his supervision. I would like to also thank him for his guidance and the discussions through my project, as well as for his flexibility in adapting my research due to the Covid-19 pandemic.

I would also like to thank Dr. Christian Mata for introducing me into the Biomedical Image Processing research field and providing support and guidance during my dissertation.

I’m grateful to my cousin Dr. Alba Gonzalez-Franquesa for the help she has provided, and the tools she has shown me in order to structure and elaborate my project. Her scientific knowledge has helped me to develop and complete this project.

I would like to thank my family for their emotional support and for encouraging me during my university career. A special thank you to dad, Anna and Carles for being close despite the distance.
Abstract

Breast cancer a global health challenge as it is the most common cancer diagnosed in women. An early diagnosis can increment the survival rates, thus detection programs aim to obtain mammography images of thousands of asymptomatic women.

Nowadays, radiologists study and describe a mammography using a standard system known as Breast Imaging Reporting and Data System (BI-RADS). This system classifies images into 6 different categories depending on the pathology found on the breast. In this project, 796 mammography images have been automatically classified by applying two different algorithms: Digital Mammography Challenge algorithm (DMC) and Convolutional Neural Network algorithm (CNN).

Different accuracy have been observed by using both algorithms: DMC achieved a 78.44% of accuracy (namely correct classification) while CNN has achieved a 82.16% in the train performance and a 62.5% in the test.

Key words

Breast cancer, Mammography, Image Processing, Image Classification, Breast Density, BI-RADS, Convolutional Neural Networks.
Resumen

El cáncer de mama es un reto para la salud mundial, ya que es el cáncer más común diagnosticado en mujeres. Las tasas de supervivencia de este cáncer pueden aumentar a partir del diagnóstico precoz de la patología, por lo que se establecen programas de cribado para mujeres asintomáticas.

Hoy en día, los radiólogos y radiólogas estudian y describen las mamografías a partir de un sistema estándar conocido como Breast Imaging Reporting and Data System (BI-RADS). Este sistema clasifica las imágenes en 6 categorías diferentes según las características y/o patologías del pecho. En este proyecto, se han clasificado de manera automática 796 mamografías aplicando dos algoritmos diferentes: ”Digital Mammography Challenge” (DMC) y Convolutional Neural Network (CNN).

Después de la aplicación de ambos métodos de clasificación se han obtenido los siguientes resultados: DMC ha alcanzado un 78,44 % de precisión (es decir, clasificación correcta) mientras que CNN ha conseguido un 82,16 % en el algoritmo de entrenamiento y un 62,5 % en el test.

Palabras clave

Cáncer de mama, Mamografía, Procesamiento de imágenes, Clasificación de imágenes, Densidad mamaria, BI-RADS, Convolutional Neural Netwroks.
El càncer de mama és un repte per a la salut mundial, ja que és el càncer més comú diagnosticat en dones. Les taxes de supervivència d’aquest càncer poden augmentar a partir d’un diagnòstic precoç de la patologia, motiu pel qual s’estableixen programes de cribatge per a dones asimptomàtiques.

Avui en dia, els radiòlegs i radiòlogues estudien i descriuen les mamografies a partir d’un sistema estàndard conegut com a Breast Imaging Reporting and Data System (BI-RADS). Aquest sistema classifica les imatges en 6 categories diferents segons les característiques i/o patologies del pit. En aquest projecte, s’han classificat de manera automàtica 796 mamografies aplicant dos algoritmes diferents: “Digital Mammography Challenge” (DMC) i Convolutional Neural Network (CNN).

Després de l’aplicació d’ambdós mètodes de classificació s’han obtingut els següents resultats: DMC ha assolit un 78,44% de precisió (és a dir, classificació correcta) mentre que CNN ha aconseguit un 82,16% en l’algoritme d’entrenament i un 62,5% en el test.

Paraules clau

Càncer de mama, Mamografia, Processament d’imatges, Classificació d’imatges, Densitat mamària, BI-RADS, Convolutional Neural Netwroks.
Contents

Acknowledgements ...................................................... viii
Abstract ................................................................. x
Resumen ................................................................. xii
Resum ................................................................. xiv
List of Figures ............................................................. xvii
List of Tables ............................................................... xix
List of Listings ............................................................. xxi
List of Acronyms ........................................................... xxiii

1. Preface ................................................................. 1
   1.1 Rationale .......................................................... 1
   1.2 Motivation ........................................................ 1

2. Introduction ........................................................... 3
   2.1 Breast Cancer ...................................................... 3
   2.2 Mammogram: Screening and Diagnosis Method ................. 5
   2.3 Classification of Images Depending on the Pathological Category .......... 8
   2.4 Project Overview .................................................. 10

3. Mammography Imaging Analysis and Interpretation .................. 11
   3.1 Pre-Processing ..................................................... 11
       3.1.1 Image Enhancement by Noise Reduction and Removal ............ 11
       3.1.2 Breast Segmentation .......................................... 12
   3.2 CAD Systems: Computerized Approaches to Mammography Analysis ...... 13
   3.3 Digital Mammography Challenge (DMC) Method Outline .............. 15
   3.4 DMC Classification Algorithm: Feature Extraction by Texture Analysis ... 15
       3.4.1 Gray Level Co-occurrence Matrix ............................. 16
       3.4.2 Laws’ Texture .................................................. 18
       3.4.3 Local Binary Patterns ........................................ 19
3.5 Convolutional Neural Networks (CNN) Method Outline .................................................. 20
3.6 CNN as a New Mammography Analysis Tool ................................................................. 20
3.6.1 AlexNet Applied in Pathological Classification of Mammography Images ...................... 23

4. Aims and Hypothesis .............................................................................................................. 27

5. Materials and Methods ........................................................................................................ 29
5.1 Mammography Data Collection ......................................................................................... 29
5.2 Mammography Images Pre-Processing ............................................................................... 30
5.2.1 Noise Reduction and Removal ....................................................................................... 30
5.2.2 Automatic Breast Segmentation ..................................................................................... 31
5.3 DMC Algorithm .................................................................................................................. 33
5.3.1 GLCM ............................................................................................................................ 33
5.3.2 Laws’ Masks ................................................................................................................... 37
5.3.3 LBP ................................................................................................................................. 39
5.4 AlexNet Algorithm .............................................................................................................. 41
5.5 Analysis of the Results: Confusion Matrix ......................................................................... 41
5.5.1 Kappa (k) Coefficient for Confusion Matrices ................................................................. 42

6. Results .................................................................................................................................. 45
6.1 Evaluation of Pre-Processed Mammography Images ......................................................... 45
6.2 Evaluation of Breast Composition Classification .............................................................. 47
6.3 Summary of Classification Results ...................................................................................... 50

7. Discussion ............................................................................................................................. 51

8. Conclusion ............................................................................................................................. 55

9. Reflections ............................................................................................................................. 57

Bibliography ............................................................................................................................ 66
List of Figures

2.1 Statistics of Cancers Cases and Deaths Worldwide . . . . . . . . . . . . . . 3
2.2 Breast Cancer Incidence and Mortality Rates Worldwide . . . . . . . . . . 4
2.3 Breast Cancer Impact in European Women . . . . . . . . . . . . . . . . . . 4
2.4 Mammogram Illustration . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
2.5 Mammography Standard Views . . . . . . . . . . . . . . . . . . . . . . . . . 7
2.6 Artefacts in Mammography Images . . . . . . . . . . . . . . . . . . . . . . 7
2.7 Adapted BI-RADS Classification . . . . . . . . . . . . . . . . . . . . . . . 10
2.8 Project Method Diagram . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10
3.1 Step-by-Step Breast Segmentation . . . . . . . . . . . . . . . . . . . . . . . 12
3.2 Spiculated Masses . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13
3.3 False Diagnoses Proportion . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
3.4 DMC Method Outline . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
3.5 Direction and Distance in a GLCM . . . . . . . . . . . . . . . . . . . . . . 16
3.6 Co-occurrence Matrix Generation . . . . . . . . . . . . . . . . . . . . . . . 16
3.7 Statistics of GLCM . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
3.8 Laws’ Statistics Descriptors . . . . . . . . . . . . . . . . . . . . . . . . . . 18
3.9 LBP computation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
3.10 Basic LBP Operator . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
3.11 CNN Method Outline . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
3.12 AI Hierarchy . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
3.13 CNN Architecture . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
3.14 AlexNet Architecture . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
3.15 Max Pooling Simplified Example . . . . . . . . . . . . . . . . . . . . . . . 24
3.16 Stride Simplified Example . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
4.1 Project Context . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27
5.1 Representative Result of Segmentation . . . . . . . . . . . . . . . . . . . . 32
5.2 Re-scaled Image in GLCM Code . . . . . . . . . . . . . . . . . . . . . . . . 34
5.3 Boxplot of Statistics in GLCM . . . . . . . . . . . . . . . . . . . . . . . . . 35
5.4 Matlab Data Tips of Proportion in Images . . . . . . . . . . . . . . . . . . 36
5.5 GLCM Classification . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37
5.6 Laws’ Masks Result . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38
5.7 LBP Histogram Result . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39
5.8 Efficient LBP Image and Pixel-wise LBP Image Result . . . . . . . . . . . 40
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.9</td>
<td>AlexNet Architecture Performed</td>
<td>41</td>
</tr>
<tr>
<td>5.10</td>
<td>Conditions in a Confusion Matrix</td>
<td>42</td>
</tr>
<tr>
<td>6.1</td>
<td>Representative Filtered-Segmented BI-RADS I Image</td>
<td>45</td>
</tr>
<tr>
<td>6.2</td>
<td>Representative Filtered-Segmented BI-RADS II Image</td>
<td>46</td>
</tr>
<tr>
<td>6.3</td>
<td>Representative Filtered-Segmented BI-RADS III Image</td>
<td>46</td>
</tr>
<tr>
<td>6.4</td>
<td>Representative Filtered-Segmented BI-RADS IV Image</td>
<td>46</td>
</tr>
<tr>
<td>6.5</td>
<td>GLCM Confusion Matrix</td>
<td>47</td>
</tr>
<tr>
<td>6.6</td>
<td>AlexNet train Confusion Matrix</td>
<td>48</td>
</tr>
<tr>
<td>6.7</td>
<td>AlexNet test Confusion Matrix</td>
<td>49</td>
</tr>
<tr>
<td>7.1</td>
<td>Representative Incorrect Segmented Images</td>
<td>52</td>
</tr>
</tbody>
</table>
List of Tables

2.1 BI-RADS Classification .............................................. 9
3.1 Laws’ Masks ............................................................... 18
3.2 State-of-art of Automated Classification Systems ...................... 22
3.3 Description of the AlexNet Architecture ................................. 24
5.1 Mammography Images Databases ....................................... 29
5.2 Database of this Project .................................................. 30
5.3 Established Proportion Limits .......................................... 36
5.4 Coefficient Kappa Interpretation ...................................... 43
5.5 Understanding Coefficient Kappa ...................................... 43
6.1 Summary of Classification Results .................................... 50
7.1 Summary of the Agreement Among Raters ............................... 53
List of Listings

5.1 Matlab Code: Copyright Section ........................................... 30
5.2 Matlab Code: GLCM Obtaining .................................................. 33
5.3 Matlab Code: Image Re-scaling ................................................ 33
5.4 Matlab Code: Statistical Descriptors Obtaining .............................. 34
5.5 Matlab Code: Classification from Proportion Limits .......................... 36
5.6 Matlab Code: Laws’ Filters and Masks Description .......................... 37
5.7 Matlab Code: Laws’ Masks Obtaining ......................................... 38
5.8 Matlab Code: LBP Histograms Obtaining ...................................... 39
5.9 Matlab Code: LBP Images Obtaining ......................................... 40
6.1 Matlab Code: GLCM Confusion Matrix and Coefficient Kappa Obtaining .. 47
6.2 Matlab Code: AlexNet train Coefficient Kappa Obtaining ................ 48
6.3 Matlab Code: AlexNet test Coefficient Kappa Obtaining ................ 50
List of Acronyms

ACR American College of Radiology
AC  American College of Surgeons
AI  Artificial Intelligence
AMA American Medical Association

BI-RADS Breast Imaging-Reporting and Data System

CAD Computer-aided Detection System
CADx Computer-aided Diagnosis
CALMa Computer Assisted Library for MAmmography
CAP College of American Pathologists
CC Cranial-Caudal
CDCP Centers for Disease Control and Prevention
CNN Convolutional NeuralNetwork

Conv Layer Convolutional Layer

DDSM Digital Database for Screening Mammography
DMC DigitalMammography Challenge

FC Fully-Connected
FDA Food and Drug Administration

GLCM Gray-Level co-occurrence Matrix
IARC International Agency for Research on Cancer
ILSVRC ImageNet Large Scale Visual Recognition Challenge
LBP Local Binary Pattern
LLNL Lawrence Livermore National Laboratory
LM Latero-Medial

MIAS Mammographic Image Analysis Society Digital Mammmogram Database
ML Medio-Lateral
MLO Mediolateral-○Oblique
MRI Magnetic Resonance Imaging
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCI</td>
<td>National Cancer Institute</td>
</tr>
<tr>
<td>ReLu</td>
<td>Rectified Linear Unit</td>
</tr>
<tr>
<td>ROI</td>
<td>Region Of Interest</td>
</tr>
<tr>
<td>STEM</td>
<td>Science, technology, engineering and mathematics</td>
</tr>
<tr>
<td>UCSF</td>
<td>University of California San Francisco</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
Preface

1.1 Rationale

Last semester, I worked on a project entitled “Digital Mammography Challenge” that consisted on the automatization of classification of 30 mammography images according to breast tissue diagnosis. That project gave me the opportunity to research and immerse myself into the biomedical image processing field, which is emerging and constantly changing.

I then gained knowledge and increased my awareness of the importance of the preprocessing, segmentation and classification systems of mammography images, all important and critical steps in analysing an ever-increasing amount of these images. Moreover, these systems would represent a great help to radiologists and ease both the first evaluation and the later diagnosis. The biggest challenge of creating an algorithm was the correct classification of any given image (namely a generic algorithm). Therefore, the Digital Mammography Challenge (DMC) algorithm I developed was based on general and common elements in all the images and it led to a significant percentage of error in the classification.

Altogether, the Master Thesis presented here aims to apply a Convolutional Neural Network (CNN) to mammography images to classify them into pathological categories and obtain a higher percentage of success in comparison with the traditional method.

1.2 Motivation

Nowadays, breast cancer is the most common type of cancer amongst women. Breast cancer has a large physiological, psychological and social impact. Its early detection and diagnosis reduce both its morbidity and mortality, making treatment decisions easier and more successful. The most commonly used technique to detect breast cancer are
mammography screenings.

The motivation for doing this research work is therefore to create a better generic algorithm that classifies mammography images by using a CNN. The results will offer a more accurate classification of the presence or absence of a tumour to the specialist.
Chapter 2

Introduction

2.1 Breast Cancer

Breast cancer is the second most common type of cancer in the world, and the most common type of cancer in women [1] (Figure 2.1). Medical research has led to enhance human life quality and expectancy. Research in oncology has developed extremely in the last decades and helped prevent, diagnose and treat malignant tumours. Hence, breast cancer represents a big challenge to oncologists around the world [2].

![Estimated number of incident cases and deaths worldwide, both sexes, all ages](image)

**Fig. 2.1:** Estimated number of incident cases and deaths worldwide, both sexes, all ages [3]

A study by Bray et al showed 2,1 million newly diagnosed cases of breast cancer in women worldwide every year [4], which represents one new diagnosed case every 18 seconds and 626,679 annual deaths [5]. There is therefore a clear need to increase the early detection and knowledge of and find a treatment for this type of cancer.
Breast cancer survival rates vary around the globe, and ranges from around 80% in North America, Japan and Sweden, to over 60% in developed countries and to less than 40% in developing countries [6] (Figure 2.2). The higher incidence of breast cancer in developed countries might be a consequence of the increased access to healthcare and therefore more number of tests and diagnosis.

![Fig. 2.2: Estimated age-standardized incidence and mortality rates (World) in 2018, all cancers, both sexes, all ages per 100,000. Graph shows 5 countries with higher and 5 countries with lower incidence [3]](image)

In 2018 the probability of a European woman being diagnosed with breast cancer was 12.5% or 1 in 8 before the age of 85 (Figure 2.3) [7]. In Europe, women under the age of 50 represents the 20% of all breast cancer diagnosed, a 36% occurs at age between 50-64, and the remaining 44% of the cases are detected in women older than these ages [7].

![Fig. 2.3: A European woman’s chance of developing breast cancer](image)
The global incidence of breast cancer is increasing annually 3.1% and likely to keep this trend [5]. Concretely, it increased from 641,000 to more than 1.6 million cases in 30 years [5]. The total number of breast cancer cases each year is expected to reach 2.4 million worldwide by 2030 [8]. Of all metastasis developed from breast cancer in different tissues, about 90% are deadly-metastasis [9]. Early detection of breast cancer reduces the probability of metastasis and thus screening (namely mammography images) in every two years is recommended in women between 40 and 74 years old [10]. This project focuses on the automatically extraction of information and classification of the images obtained by the above-mentioned mammography images.

2.2 Mammogram: Screening and Diagnosis Method

In order to reduce the mortality and the impact of breast cancer on the population several strategies have been assessed and implemented. Among them, screenings aid the early detection of primary and metastatic tumors in asymptomatic and cancer-recovered women, respectively [11, 12].

Mammograms are the most commonly performed screenings for breast cancer, along with a physical breast exploration or a breast self-examination [13, 14]. Breast magnetic resonance imaging (MRI) can complement the clinical data and help confirm the suspicion of a malignant tumor in women [15].

A mammogram is the examination by low-energy X-rays of human breast tissues. A mammography is a 2D projection of the breast tissue [16]. Images are collected for the radiologists to detect potential tumors with the aid of image processing systems. A mammogram reveals alterations in the breast by measuring its density and fat content [13]. The mammogram machine (Figure 2.4) is composed of two plates that compress the breast tissue allowing for the use of less X-ray ionization and providing better images [17].

![Mammogram illustration](image)

Fig. 2.4: Mammogram illustration. Stages to obtain a mammography (adapted from [18])
Chapter 2. Introduction

Due to the strong compression on the breast and other parameters such as the test position, mammographic screenings can present, apart from the exposure to radiation, anxiety in the patient, false positive results and overdiagnosis [15]. Preston et al studied radiation-induced cancer and realized a higher risk of developing breast cancer with increased levels of radiation, which diminishes with an increasing age [19]. Indeed, women under the age of 20 show a higher risk of radiation-based cancer [20], after a mammogram, leading to a radiosensitive breast tissue [19]. Therefore, the damage-effectivity of these screening tests is being constantly studied.

During 1980s, a number of clinical trials demonstrated the efficacy of mammography screening [12, 21] reason why they were implemented in richer countries and became nationally recommended in developing and poor countries, together with physical breast examination and breast self-examination [15]. In 2002, the International Agency for Research on Cancer (IARC) concluded that breast cancer death reduced from 20% to 25% due to mammography screening of women aged 50 to 74 [22]. In line with this, Chetlen et al showed in 2015 that breast cancer mortality was reduced by 20% in all ages due to mammographic screening [23]. This study confirmed the benefit of mammograms as screenings and encouraged to improve and invest in new screening modalities to make an early diagnosis to patients on whom mammogram was insufficient [23]. In the same year, IARC revised the results of randomized controlled trials to reaffirm that mammographic screenings were only effective in women aged 50-69 [15].

Although mammograms are aimed to reduce the mortality in breast cancer, later studies in Europe revealed no changes in deaths in relation to the mammographic screening [24, 25, 26]. WHO described for the first time in 1968 the criteria that a patient has to follow to undergo a mammography as a screening method [27]. Independently of the current controversy of the benefit of this screening, these reasons are valid in the present. These criteria are method [27]:

- The screened disease is serious and prevalent.
- The test is sensitive and specific.
- The test is well tolerated, inexpensive and it changes therapy or outcome.

Besides, mammogram examination is the cheapest and most efficient test to detect breast cancer [28]. Collectively, mammograms are still a test of choice for a cancer diagnosis. For this reason, the growing biomedical engineering field is now focused on the obtaining of improved images [29].

These images cover two planes of the breast: oblique or angled (mediolateral-oblique, MLO) and upper (cranial-caudal view, CC) [17] (Figure 2.5). Besides, there are lots of supplementary views that aid in diagnosis when needed. Between all the additional
2.2. Mammogram: Screening and Diagnosis Method

views, two most important are (considered standards): Medio-Lateral view (ML) and Latero-Medial view (LM). This project focuses on the MLO images [30].

![Mammography standard screening views. Medio-Lateral oblique (MLO) and Cranio-Caudal (CC) views (adapted from [17, 31])](image)

Mammography images often show undesired objects or tissues that might interfere with the analysis.

Therefore, an image pre-processing is performed. In this pre-processing step, all the distortions from images are removed (Figure 2.6). These distortions appear in different forms like high intensity in some regions (e.g., labels or scanner effects), image mismatch that appears during the scanning process, presence of the pectoral muscle and others. After the pre-processing, images are segmented to include only the region of interest (ROI) (namely breast tissue) [32].

![Common artefacts occurred in mammography images [32]](image)

Once the images have been pre-processed: filtered and segmented, mammography images can be analyzed and classified.
2.3 Classification of Images Depending on the Pathological Category

Breast Imaging Report and Data System (BI-RADS) is a standard classification method based on breast pathology obtained from mammography images, ultrasound and magnetic resonance imaging (MRI) [33]. The two algorithms included in this research work aim to classify the images based on the BI-RAD System. BI-RADS is used by radiologists to interpret and communicate the results of mammograms in a standardized way in order to detect and diagnose potential breast cancer [34, 35]. BI-RADS was implemented for the first time in 1993 [36]. It is a collaborative work between the American College of Radiology (ACR), the National Cancer Institute (NCI), the Centers for Disease Control and Prevention (CDCP), the Food and Drug Administration (FDA), the American Medical Association (AMA), the American College of Surgeons (ACS) and the College of American Pathologists (CAP). The BI-RADS Committee includes members from all the aforementioned institutions [37].

BI-RADS uses breast density to classify the mammography images [38]. Breast density is the ratio of fibroglandular to fatty tissue in the breast [39]. BI-RADS lexicon also defines mass, calcification, architectural distortions, cutaneous lesions and others mammographic features [35], altogether helping radiologist to [35, 40]:

- Predict benign and/or malignant abnormalities in the breast.
- Collect data.
- Homogenize the results.
- Standardize a single guide so the classification of the results is understood by the radiologists and doctors.
- Simplify and analyze the mammography images and reduce errors in diagnosis.

To date, there are 6 different BI-RADS categories that define the risk of presenting breast cancer and the diagnosis [41] (Table 2.1):

Given that this project will only include healthy breast tissue, and with the aim to simplify the categorization process, only 4 BI-RADS categories (Figure 2.7) will be used henceforth:

- **BI-RADS I**: The breast is almost entirely fatty.
- **BI-RADS II**: There is some fibro-glandular tissue.
- **BI-RADS III**: The breast is heterogeneously.
- **BI-RADS IV**: The breast is extremely dense.
### Tab. 2.1: BI-RADS Categories, Assessment and Management

<table>
<thead>
<tr>
<th>BI-RADS Assessment</th>
<th>Definitions</th>
<th>Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 0</td>
<td>Incomplete – Additional Imaging Evaluation Required</td>
<td>Recall for additional imaging and/or comparison with prior examination(s)</td>
</tr>
<tr>
<td>Category 1</td>
<td>Negative</td>
<td>Routine mammography screening</td>
</tr>
<tr>
<td>Category 2</td>
<td>Benign</td>
<td>Routine mammography screening</td>
</tr>
<tr>
<td>Category 3</td>
<td>Probably Benign</td>
<td>Short-interval (6-month) follow-up or continued surveillance mammography</td>
</tr>
<tr>
<td>Category 4A</td>
<td>Low Malignancy Suspicious</td>
<td></td>
</tr>
<tr>
<td>Category 4B</td>
<td>Moderate Suspicion for Malignancy</td>
<td>Tissue diagnosis</td>
</tr>
<tr>
<td>Category 4C</td>
<td>High Suspicion for Malignancy</td>
<td></td>
</tr>
<tr>
<td>Category 5</td>
<td>Highly Suggestive of Malignancy</td>
<td></td>
</tr>
<tr>
<td>Category 6</td>
<td>Known Biopsy-Proven Malignancy</td>
<td>Surgical excision when clinically appropriate</td>
</tr>
</tbody>
</table>
Chapter 2. Introduction

2.4 Project Overview

This Master Thesis aims to classify mammography images depending on breast tissue features. The project has been divided into 6 different steps (Figure 2.8).

Fig. 2.7: Adapted BI-RADS classification performed in this project

Fig. 2.8: Integral block diagram for the automated classification of mammography images through two different methods
Mammography Imaging Analysis and Interpretation

Mammography imaging study includes image pre-processing and analysis, the latter consisting of digital mammography challenge (DMC), adapted from computer aided diagnosis (CAD), or convolutional neural networks (CNN). The theory behind these methods is described below.

3.1 Pre-Processing

Image pre-processing techniques are essential to reduce and remove noise in mammography images that affects the accuracy of diagnosis by generating a high false positive and negative rate [42, 43]. Thus, the main goal of this process is to improve the quality of the image and ease the diagnosis [44]. Image processing is divided into image enhancement (noise removal) and breast part extraction (segmentation) [42, 45].

3.1.1 Image Enhancement by Noise Reduction and Removal

Noise is a random and undesirable artifact produced during image acquisition or transmission [46, 47]. Noise alters the intensity of the image leading to false information [42], and it can be divided into different types: impulse, amplifier, multiplicative, poisson (shot), quantization, film grain, on-isotropic, and periodic noise [46]. The four first are the most common type of noise detected in mammography images. All of them directly affect to the entire image processing and diagnosis steps [48]. To describe image degradation during acquisition, noisy images are modeled as [48]:

\[ g(x, y) = f(x, y) * n(x, y) \]
Where $g(x, y)$ represents the degraded image, $f(x, y)$ is the original image and $n(x, y)$ is the noise. The operator $*$ makes a convolution [48]. Image denoising is an essential step that aims to recover the original expression of a given image, by the detection and the removal of all the noise on it [49]. The techniques to end with this undesirable artifact normally manipulate the above-mentioned equation in order to obtain an estimate of $f(x,y)$ of the input [48]. Depending on the noise detected in a mammography, different filters are applied.

### 3.1.2 Breast Segmentation

After noise removal, another step aims to simplify the image in order to make its processing easier: image segmentation [50]. To analyze breast tissue and its abnormalities, images are segmented into different parts (Figure 3.1): breast region (provides information to the study) and the rest (background and pectoral muscle) [50, 51, 52]. Breast segmentation is achieved in three main steps: label removal, breast localization and pectoral muscle removal [53]. Indeed, removal of muscle is crucial in MLO images as this tissue appears with a higher brightness than the rest of the image. Failure to remove the muscle would alter and provide incorrect results [52].

![Fig. 3.1: Breast region segmentation steps](image)

Segmentation is performed by obtaining a binary mask with the same dimensionality of the original image. Binarization of an image reduces the information in the image by making its values true or false (0 or 1). This is represented by the intensity range of a pixel, that can be totally black (grey value 0) or totally white (grey value 255) [42].

A part from manual segmentation [16], the most common techniques to segment the breast on a mammography [52] are based on: gradient, polynomial modelling, active contours, classifiers, or histogram. Histogram-based segmentation is the technique of choice for this project.
3.2 CAD Systems: Computerized Approaches to Mammography Analysis

The process of detecting lesions on mammography is quite a monotonous and time-consuming task for radiologists [54]. In fact, only 3 or 4 cases out of a thousand analyzed present cancerous indications, thus it is easy for an abnormality to be overlooked by them [55]. In order to solve this issue and speed the image analysis, automated computer-aided detection systems (CAD) were suggested for the first time in 1967 [56] to assist radiologists on mammography interpretation [54, 57].

CAD systems were developed to process and analyze images with the aim of labeling images and classifying tumors between benign or malign [57] through traditional texture analysis or through localizing lesions (by detection, segmentation and classification) [58]. CAD systems originally aimed to reduce the number of incorrect diagnoses by radiologists [59]. These systems alert the radiologist upon a confusing area of a medical image [60] and therefore ease the image interpretation and diagnosis by the professionals [55]. In fact, radiologists detect only 84% of breast cancers upon interpretation and CAD systems aim to mimic what these radiologists do, thus not improving the 16% of cancer left undetected because of human eye limitation [59]. Indeed, the limitation comes from various factors such as the level of experience, the perception and the sight quality of the radiologist, the image quality, the breast density, among others [61].

Radiologists take into account the shape of breast tissue, the edges (circumscribed masses or not), the density of the mass in relation to the rest of the breast, the association with other abnormalities (asymmetries, architectural distortions, among others) and the comparison between different screenings to generate a holistic diagnostic approach [62, 63]. Of all these features, the analysis of the edges and density of the masses found on a mammography are essential for the diagnosis of the breast cancer. The most common type of malignant mass in breast cancer are spiculated masses [55, 64], representing the 96% of malignant breast tumors [63]. Spiculated masses are characterized by a highly dense center which radiates lines or spicules to its margin (Figure 3.2) [55, 65].

![Fig. 3.2: Different masses in breast. (a) Benign circumscribed mass ; (b) Malignant circumscribed mass ; (c) Malignant spiculated mass ; (d) Malignant spiculated mass](image-url)
CAD systems detect better microcalcification than spiculated masses \cite{55, 66}. As calcifications are denser than surrounding tissue, it is therefore easier for CAD systems to detect and differentiate calcifications from the breast tissue due to the high contrast between them. Instead, spiculated masses present a wide variability of physical characteristics that make it harder to standardize \cite{66}. This variability of the characteristics of the masses alters the orientation of the diagnosis making it more complex and therefore leading to errors on it. False positives and negatives are a current problem in breast cancer diagnosis through images \cite{67} leading to an analysis of a mammography that indicates breast cancer when in fact there is none \cite{1} or that indicates there is no breast cancer when there is. On one hand, false positives represent approximately 10%-20% of the patient diagnosis \cite{1, 15}, out of which 90% need complementary tests such as a biopsy \cite{1, 68}. False positives in breast cancer have been demonstrated to cause a negative psychological consequence in incorrectly-diagnosed patients \cite{15, 69}. On the other hand, false negatives represent 15%-20% of the diagnosis and are a limitation for screening, most commonly in dense breasts \cite{1}. False negatives involve complementary tests like exploration of a palpable mass or suspicious shape of the nipple \cite{1, 57}. The huge impact of false diagnosis and the importance of early detection of breast cancer are current limitations on CAD systems(Figure 3.3).

To date, automatic classification of mammography images is an unsolved problem \cite{70}. Deep Learning methods for mammography images such as Convolutional Neural Networks (CNN) \cite{71} are leading to new methods for image analysis.

**Fig. 3.3:** False positive (10% - 20%) and negative (15% - 20%) diagnoses proportion compared to total diagnoses
3.3 Digital Mammography Challenge (DMC) Method Outline

DMC has been inspired on the traditional automated system to classify mammography images according to 4 adapted BI-RADS. In order to achieve this goal, DMC algorithm has focused into the analysis of different texture features through Gray-Level co-occurrence Matrix (GLCM), Laws’ texture Masks and Local Binary Patterns (LBP) (Figure 3.4).

![Block diagram of the DMC method](image)

**Fig. 3.4:** Block diagram of the DMC method

3.4 DMC Classification Algorithm: Feature Extraction by Texture Analysis

After noise removal and segmentation, features are extracted for classification of breast tissue according to the 4 predefined BI-RADS (detailed in 2.3). Features are extracted to describe visual patterns and shapes in the breast through quantitative values [72], used for the automation of image classification. This can be performed by different methods: texture analysis, intensity distribution or learning from images through artificial intelligence (AI) [39]. Texture analysis presents many medical applications and it is the most commonly used in mammography analysis [72, 73, 74]. In a given image, texture is defined as the repetition of intensity patterns [75]. Texture plays a key role in the detection of masses [73], as it contains information about the intern structure of breast tissues [76]. This method has proven successful in differentiating tissues for breast cancer detection [73, 74].
However, texture representation in two-dimensional (2-D) is a challenging problem as it presents variations in orientation or scale given the lack of uniformity in the three-dimensional (3-D) [72, 74].

To date, there are several studies that have succeed in texture extraction. In this project three different texture descriptors are evaluated in order to describe the texture in mammographic images: Grey Level Co-occurrence Matrix (GLCM), Law image filter and Local Binary Pattern (LBP).

### 3.4.1 Gray Level Co-occurrence Matrix

As the name suggests, a GLCM is a matrix that defines the gray-level frequency in a given image. It shows the result of counting the number of times that a pair of pixels appear in the image (Figure 3.5). These 2 pixels are normally neighbors but they can also be separated by a distance “d” and angle $\theta$, that can be $0^\circ$, $45^\circ$, $90^\circ$, and $135^\circ$ [77, 78]. This matrix provides the statistical distribution of intensities (pixel value) and it creates a co-occurrence matrix (2-D histogram) [53, 79].

![Fig. 3.5: Direction and distance between pixel of interest and its neighbor](image)

The following simplified example shows how GLC matrix is the result of counting the number of times that a given pair of pixels appear in the 4x4 input matrix. In this case, the pair (2,4) appear 2 times (Figure 3.6).

![Fig. 3.6: Representative example of co-occurrence matrix generation (d=1, \theta=0)](image)
This created co-occurrence matrix allows for the extraction of different textural features depending on the distance and the direction. Thus, it is possible to obtain the following statistical features that can describe texture: angular second moment, energy, entropy, contrast, correlation, dissimilarity, sum average, sum entropy, sum variance, difference variance, difference entropy, and homogeneity features (Figure 3.7) [77]:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angular sec. moment</td>
<td>$f_1 = \sum_{i,j=0}^{N-1} P(i,j)^2$</td>
</tr>
<tr>
<td>Energy</td>
<td>$f_2 = \sqrt{f_1}$</td>
</tr>
<tr>
<td>Entropy</td>
<td>$f_3 = \sum_{i,j=0}^{N-1} P(i,j)(-\log P(i,j))$</td>
</tr>
<tr>
<td>Contrast</td>
<td>$f_4 = \sum_{i,j=0}^{N-1} P(i,j)(i-j)^2$</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>$f_5 = \sum_{i,j=0}^{N-1} \frac{1}{1+(i-j)^2}P(i,j)$</td>
</tr>
<tr>
<td>Correlation</td>
<td>$f_6 = \sum_{i,j=0}^{N-1} \frac{(i-\mu)(j-\mu)}{\sqrt{\sigma^2 \sigma^2}}$</td>
</tr>
<tr>
<td>Sum average</td>
<td>$f_7 = \sum_{k=2}^{2N} \sum_{i+j=k}^{N-1} P(i,j)$</td>
</tr>
<tr>
<td>Sum variance</td>
<td>$f_8 = \sum_{k=2}^{2N} \sum_{i+j=k}^{N-1} (k-f_7)^2 P(i,j)$</td>
</tr>
<tr>
<td>Sum entropy</td>
<td>$f_9 = -\sum_{k=2}^{2N} \sum_{i+j=k}^{N-1} P(i,j) \log(P(i,j))$</td>
</tr>
<tr>
<td>Difference variance</td>
<td>$f_{10} = \sum_{k=0}^{2N} \sum_{i+j=0}^{N-1} (k-\mu k)^2 P(i,j)$</td>
</tr>
<tr>
<td>Difference entropy</td>
<td>$f_{11} = \sum_{k=0}^{2N} \sum_{i+j=0}^{N-1} P(i,j) \log(P(i,j))$</td>
</tr>
<tr>
<td>Dissimilarity</td>
<td>$f_{12} = \sum_{i,j=0}^{N-1} P(i,j)</td>
</tr>
</tbody>
</table>

**Fig. 3.7:** Different feature statistics that can be evaluated on a co-occurrence matrix [77]

During this project, the following statistical features have been selected to compute GLCM:

- **Contrast:** measure of intensity difference between a pixel and its neighbor over the entire image. If image contrast is 0, image is constant.

- **Homogeneity:** spatial closeness of the distribution of the co-occurrence matrix.

- **Correlation:** measure of how correlated a pixel is to its neighbor over the whole image.

- **Proportion:** number of white pixels over the total number of breast pixels. It represents the tissue into the breast.
3.4.2 Laws’ Texture

In 1980, Kenneth Ivan Laws suggested a texture energy measurement known as Laws’ texture filter or Laws’ features [80]. Laws’ features consist of five one dimensional (1-D) filters that are designed to detect and obtain different type of structures in an image. These five filters are the result of the convolution of the three 1 x 3 vectors that he first proposed: L3 (averaging) = [1, 2, 1], E3 (edges) = [-1, 0, 1] and S3 (spots) = [-1, 2, -1]. Five 1-D Laws’ vectors are [81, 82]:

- E5 (edges) = [-1, -2, 0, 2, 1] = L3 * L3
- S5 (spots) = [-1, 0, 2, 0, -1] = - E3 * E3 = L3 * S3
- R5 (ripples) = [1, -4, 6, -4, 1] = S3 * S3
- W5 (waves) = [-1, 2, 0, -2, 1] = L3 * E3
- L5 (low pass or average gray value) = [1, 4, 6, 4, 1] = - E3 * S3

The result of the convolution of an horizontal 1-D filter with a vertical one is 25 different 2-D filters (known as Laws’ masks) (Table 3.1) [81, 82].

<table>
<thead>
<tr>
<th>Tab. 3.1: Laws' Masks. Result of the convolution of 1-D Laws' filters</th>
</tr>
</thead>
<tbody>
<tr>
<td>L5L5 E5L5 S5L5 W5L5 R5LR</td>
</tr>
<tr>
<td>L5E5 E5E5 S5E5 W5E5 R5E5</td>
</tr>
<tr>
<td>L5S5 E5S5 S5S5 W5S5 R5S5</td>
</tr>
<tr>
<td>L5W5 E5W5 S5W5 W5W5 R5W5</td>
</tr>
<tr>
<td>L5R5 E5R5 S5R5 W5R5 R5R5</td>
</tr>
</tbody>
</table>

Laws’ masks are convoluted with an image to obtain its statistical parameters [82]: mean, standard deviation and absolute mean (Figure 3.8) [83].

\[
\text{Mean } \mu = \frac{1}{N} \sum_{i=1}^{N} y_i \\
\text{Standard deviation } s = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \mu)^2} \\
\text{Absolute mean } absmean = \frac{\sum_{i=1}^{N} |y_i|}{N}
\]

Fig. 3.8: Laws’ statistics descriptors that can be evaluated from Laws’ masks [83]
3.4.3 Local Binary Patterns

In 1996, Ojala et al. [84] described LBP method for the first time. This method consists of thresholding the neighborhood of each pixel with the center value (Figure 3.9).

![LBP computation](image1)

**Fig. 3.9:** LBP computation. Example of the basic LBP operator [73]

Depending on whether the pixel of interest is higher, equal or lower to the center pixel, it is labeled as 1, 1 and 0, respectively [85]. The result of this thresholding method is a binary number and an LBP code that represent the local information. All the LBP codes are used to label the pixels of the image in order to obtain an histogram of the labeled image and texture descriptors [77]. After LBP is done, statistical parameters to describe breast density will be calculated: mean, variance, standard division, skewness and kurtosis [77]. The LBP code is obtained by giving 8 weights to a matrix starting from 1 to 128 (multiplying by 2 the previous number each time). LBP code is the result of summing up the values of weights pixels multiplied with the threshold matrix (Figure 3.10).

![Basic LBP Operator](image2)

**Fig. 3.10:** Basic LBP Operator. Representation of the basic performance of a LBP operator

Altogether, GLCM, Laws’ texture filter and Local Binary Pattern (LBP) are used as texture descriptors. In this project, I present the results after using GLCM for image classification, given the challenges I experienced when using the other methods.
3.5 Convolutional Neural Networks (CNN) Method Outline

Convolutional Neural Networks (CNN) has classified the images by dividing the dataset into two subsets: training set and testing set. The results of interest are the ones obtained in the testing set (Figure 3.11).

![Fig. 3.11: CNN method outline](image)

3.6 CNN as a New Mammography Analysis Tool

CNN are classified as a subgroup of Deep Learning, that is within machine learning which is a subclass of artificial intelligence (AI) (Figure 3.12) [86]. The sensitivity of diagnostic guidance from AI does not differ much from that of the radiologist, but it facilitates and shortens the time of making a diagnosis, as well as reducing the number of false positives and negatives [87]. Concretely, CNN is the most widely used Deep Learning technique [88] in which the whole image is employed as input data [86, 89].

![Fig. 3.12: AI Hierarchy. Venn diagram representation of the AI hierarchic terminology](image)
In the last few years, CNNs have been applied in the field of biomedical engineering, especially in radiology [90]. In terms of biomedical image processing, the use of this algorithm has improved considerably image recognition, segmentation and detection of bodies and regions [91]. CNN is an image processing algorithm whose layer architecture is inspired by the organization of the visual cortex and reminds the neuronal network of the human brain [88]. This architecture is a composition of different layers of mathematical functions [92] that allows the algorithm to learn features from the image data instead of using analytically extracted features as CAD systems do [86, 89]. These layers apply convolutional operations to identify the presence or absence of edges and particular orientations of the objects of the image (first layer). Detect shapes by analyzing edges (second layer). Relate different parts of an object (third layer), and finally the following layers combines these parts to detect objects (Figure 3.13) [86]. Unlike traditional Artificial Neural Networks, CNNs automatically extract features from the image to perform object classification or segmentation.

![CNN Architecture Diagram]

**Fig. 3.13:** CNN architecture. Simplified example of a CNN performance [71]

In relation to breast cancer diagnosis, CNN has advanced considerably in processing mammography [71]. CNN is mainly applied in breast cancer for detection and classification of mammography images, a goal that has yet not been successfully achieved and that is the reason why this field is still researched [86, 93]. CNN assists radiologists with precise quantitative analysis of suspicious lesions by detecting and locating abnormalities in the breast, feature extraction, image retrieval and classification of images [71]. CNN could, as the previous systems used for automatically classification of the images by breast composition, focus on the differencing between benign or malign tumors or on the classification of images into BI-RADS (as explained in sub-section 2.3) [86]. Understanding previous studies on the classification of mammographic images contextualize this project. To date, different systems used for mammography classification have shown different correct diagnosis percentages (performance), as shown in Table 3.2 [94].
### Tab. 3.2: State-of-art of automated classification systems focusing on BI-RADS [93]

<table>
<thead>
<tr>
<th>Study year</th>
<th>Aim</th>
<th>Method</th>
<th>Performance</th>
<th>Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percha, 2012 [93]</td>
<td>To classify mammography reports into BI-RADS breast tissue composition categories</td>
<td>Rule-based</td>
<td>&gt;99% classification accuracy</td>
<td>Only focus on breast tissue composition</td>
</tr>
<tr>
<td>Hussam, 2013 [28]</td>
<td>Present a BI-RADS features extraction algorithm for clinical data mining</td>
<td>Rule-based</td>
<td>97.7% precision, 95.5% recall</td>
<td>Does not classify BI-RADS final assessment categories</td>
</tr>
<tr>
<td>Sippo, 2013 [95]</td>
<td>To extract BI-RADS final assessment categories from the reports</td>
<td>Rule-based</td>
<td>99.6% precision 100%, recall</td>
<td>If BI-RADS category is not reported, it can not classify or predict the BI-RADS category</td>
</tr>
<tr>
<td>Bozkurt, 2016 [29]</td>
<td>Using automatically extracted information from mammography reports for decision-support</td>
<td>Rule-based + supervised ML</td>
<td>Accuracy=97.58%</td>
<td>Generalizability issues since the training set is from a single center and semi-structured</td>
</tr>
<tr>
<td>Castro, 2017 [2]</td>
<td>To extract BI-RADS final assessment categories from the reports</td>
<td>Rule-based + supervised ML</td>
<td>Precision=98% Recall = 93%</td>
<td>Can only extract the BI-RADS category if BI-RADS category is reported by the mammographer</td>
</tr>
<tr>
<td>Gupta, 2018 [96]</td>
<td>To extract relations in an unsupervised way from radiology reports</td>
<td>Rule-based + unsupervised clustering</td>
<td>Precision=95% Recall = 94%</td>
<td>Do not classify BI-RADS final assessment categories</td>
</tr>
</tbody>
</table>
3.6.1 AlexNet Applied in Pathological Classification of Mammography Images

AlexNet is a CNN algorithm used in this study in order to classify the images according to the 4 predefined BI-RADS categories. AlexNet was designed by Alex Krizhevsky et al [97] and was the 2012 winner of ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [97]. AlexNet was a breakthrough in the field of deep learning for image recognition and classification [98]. The architecture of this CNN algorithm is divided into 8 layers: 5 convolutional layers (Conv Layers) that work on the feature extraction and 3 fully-connected layers that work on data classification (Figure 3.14) (Table 3.3) [97, 98].

Fig. 3.14: AlexNet architecture. An illustration of the AlexNet architecture using CNN with filter values at each layer [97]

The five Conv Layers are:

1. Layer of 96 kernel filters of size 11x11 [98, 99] that are enough to capture the entire object. Kernel filter is a matrix that slides across the input image to enhance or highlight features on it.

2. Layer with filter shape reduced to 5x5 (with 256 kernels) and its followed by 3.

3. Layers with convolutional windows shape is 3x3 (with 384 kernels) [97].

4. Fourth layer has the same characteristics than layer 3.

5. Layer with 256 kernels of size 3x3.

The fully-connected layers are performed of 4096 neurons each and the output layer has the possibility to differentiate into 1000 different categories [97]. This information is summarized in the table below (Table 3.3).
Tab. 3.3: Architecture of AlexNet algorithm. Max pooling: maximum pooling. FC: fully-connected [97]

<table>
<thead>
<tr>
<th>Layer</th>
<th>Feature Map</th>
<th>Size</th>
<th>Kernel Size</th>
<th>Activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>Image</td>
<td>1 227x227x3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>Convolution</td>
<td>96 55x55x96</td>
<td>11x11</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Max pooling</td>
<td>96 27x27x96</td>
<td>3x3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Convolution</td>
<td>256 27x27x256</td>
<td>5x5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Max pooling</td>
<td>256 13x13x256</td>
<td>3x3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Convolution</td>
<td>384 13x13x384</td>
<td>3x3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Convolution</td>
<td>384 13x13x384</td>
<td>3x3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Convolution</td>
<td>256 13x13x256</td>
<td>3x3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Max pooling</td>
<td>256 6x6x256</td>
<td>3x3</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>FC</td>
<td>- 9216</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>FC</td>
<td>- 4096</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>FC</td>
<td>- 4096</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Output</td>
<td>FC</td>
<td>- 1000</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

These layers can include several steps:

- **Max pooling**: aims to simplify the input image by reducing its dimensions (width and height) in order to help the over-fitting, a challenging step in Neural Network analysis for the new data to fit and behave like the trained data [100]. An abstracted form of the image is obtained after applying a max pooling layer, that uses a “max filter” to reduce the number of parameters on the image ultimately reducing the data computation (Figure 3.15) [100].

![Fig. 3.15: Max pooling simplified example. Max pool with 2x2 filters and stride 2](image-url)
3.6. CNN as a New Mammography Analysis Tool

- **Stride**: controls how filter convulses around the input image. It is the distance between the centers in every convolution of a kernel filter (Figure 3.16) [97].

![Stride example](image)

**Fig. 3.16**: Stride simplified example. First case (green) shows the convolution of a 3x3 kernel filter with a 7x7 matrix and stride 1 generating a 5x5 output. Second case (blue), conditions are the same except stride that changes to 2. Output is a 3x3 matrix.

- Fully-Connected (FC) layers take the output from the previous layer and classifies the output into a label (simple classification) [101].

- Rectified Linear Unit (ReLu): the output of a neural network is determined by a mathematical activation function, ReLu layer applies \( f(x) = \max(0, x) \) function to all the values in the input image in order to change all the negative activation to 0 [102]. It has been shown that deep CNNs work faster using ReLu than when they are trained with the saturating activation functions (tanh or sigmoid) [100].

Other important parameter to consider in the layers is:

- **Feature map**: is the number of filters applied in the previous layer. Each filter corresponds to a small part of the image that matches with a feature [103].
Aims and Hypothesis

Aims

I To apply a Convolutional Neural Network to mammography images in order to classify them into different density categories.

II To determine if the application of Convolutional Neural Network allows for a better density classification in comparison to the Digital Mammography Challenge algorithm.

Hypothesis

The use of AlexNet algorithm allows to obtain a higher number of correct classifications of mammography images into different density categories, thus diminishing the number of misclassified samples.
Materials and Methods

5.1 Mammography Data Collection

Obtaining and working with mammography images can become a challenge given the limited amount of available data [104]. These databases are crucial for the professionals that work with this type of images daily [105]. Specially for amateur radiologists that need to be trained. It is important that these databases include both normal breast and breast with abnormalities [105]. To date, the scientific literature contains public and private databases [104], among them: MIAS (Mammographic Image Analysis Society Digital Mammogram Database) [106], DDSM (Digital Database for Screening Mammography) [107], LLNL/UCSF database (Lawrence Livermore National Laboratory / University of California, San Francisco) and CALMa (Computer Assisted Library for MAmmography) [108] (Table 5.1).

Tab. 5.1: Mammography Images Databases. Four of the most popular mammography images databases

<table>
<thead>
<tr>
<th></th>
<th>MIAS</th>
<th>DDSM</th>
<th>UCSF/LLNL</th>
<th>CALMa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin</td>
<td>UK</td>
<td>USA</td>
<td>USA</td>
<td>Italy</td>
</tr>
<tr>
<td>Number of images</td>
<td>322</td>
<td>10480</td>
<td>198</td>
<td>3000</td>
</tr>
<tr>
<td>File Access</td>
<td>Free</td>
<td>Free</td>
<td>Paid ($100)</td>
<td>Closed</td>
</tr>
<tr>
<td>Image File Type</td>
<td>PGM</td>
<td>LJPE</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>G/JPEG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For this project, 796 DDSM images (Mamo 1) and 322 MIAS (Mamo 2) have been used. All the images are in MLO view and have been converted into Portable Network Graphics (PNG) format with 8-bit depth gray scale pixels. The intensity range of pixels comprises from 0 to 255 (from black to white). All the images are from healthy breasts,
so none of them presents abnormalities (namely tumors). Each image is named as _b1, _b2, _b3 or _b4 depending on the BI-RADS category (Table 5.2):

**Tab. 5.2:** Database of this project. Mamo 1 contains 796 images and Mamo 2 group 322 images

<table>
<thead>
<tr>
<th>BI-RADS</th>
<th>Mamo 1</th>
<th>Mamo 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (b1)</td>
<td>196</td>
<td>86</td>
</tr>
<tr>
<td>II (b2)</td>
<td>200</td>
<td>103</td>
</tr>
<tr>
<td>III (b3)</td>
<td>200</td>
<td>96</td>
</tr>
<tr>
<td>IV (b4)</td>
<td>200</td>
<td>37</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>796</strong></td>
<td><strong>322</strong></td>
</tr>
</tbody>
</table>

Initially, only 400 images from Mamo 1 were used to train AlexNet CNN, which were not sufficient to obtain acceptable results, thus the number of images was increased. Concretely, 396 images were added to Mamo 1, and Mamo 2 was created with 322 new images and also added to the project. Ultimately, as Mamo 1 folder included images with different resolution compared to Mamo 2, this project has just used images from Mamo 1 to work. Altogether, the total of images used in this project have been 796 (196 of BI-RADS 1 and 200 of each other BI-RADS).

## 5.2 Mammography Images Pre-Processing

This section includes noise removal and segmentation techniques performed in the project. However, the code executed can not be shown and shared as it has been adapted from a copyright code. The original code is a restricted material for own use that is in normal practice and rights are protected:

**Listing 5.1:** Matlab Code: Copyright Section

```
*********COPYRIGHT segmentation ************
```

All the functions needed from image preprocessing until DMC classification have been implemented with Matlab (Matlab R2020a version). For the processing of the images, several add-ons were installed with license separately: Image Processing Toolbox and Wavelet Toolbox.

### 5.2.1 Noise Reduction and Removal

In this first step of mammography processing, two types of filters ha been applied to reduce and remove noise from images. As explained in the section 2.2, mammography
images often present same type of noise: impulse (salt-and-pepper), amplifier (gaussian), multiplicative (speckle) and poisson (shot). After confirming the presence of these noises in most of the images, Gaussian and morphological filters have been applied.

Gaussian filter is a low pass filter based on peak detection, assuming that peaks are to be impulses. The spectral coefficient of interest and the amplitude spectrum coefficient are corrected by this filter within the filter window. The Gaussian filter convolutes each pixel of the input matrix with a Gaussian kernel and sums them to produce the output matrix (image without noise).

Morphological filters – dilate, erode, open and close – are non-linear operations related to the shape of different objects that constitute an image. Morphological filters have been applied in order to delete salt-and-pepper noise and to make a preliminary removal of tags in the image. This removal has been achieved by labeling all the objects in the mammography and deleting the ones which are smaller than the breast. Given that this step have not achieved a perfect delimitation of the breast from the background, this has been completed by a segmentation step.

5.2.2 Automatic Breast Segmentation

In this project, segmentation was initially proposed to be based on the Matlab-based open software: “OpenBreast: Computerized Analysis of Mammography Images” by Said Pertuz et al. [53]. After trying this open software and non-succeeding due to some problems with the images format, the segmentation part has been performed with an adaptation of the code developed by Dr. Christian Mata in the projects (currently protected): “Descriptors applied to Digital Mammography” last 2009 [83], and “Web-based application for medical imaging management” last 2015 [109]. The adaptation of these protected code have segmented the mammography images through two clearly differentiated steps:

1. **Automatic thresholding algorithm**: have been focused on segmenting the breast from the background, thus removing labels and tape artifacts from that area.

2. **Region growing algorithm**: have separated the region of interest (breast) from the pectoral muscle trying not to remove any part from breast tissue.

In order to develop this segmentation code, several steps have been taken:

1. To construct an intensity histogram of the image.

2. To use a threshold to differentiate breast and pectoral muscle from the background through an algorithm that recovers the interesting differentiated part (breast and pectoral muscle).
3. To extract the pectoral muscle region by generating another histogram of this new extracted part prior to application of the region growing algorithm.

4. To use morphological operations to smooth breast edges.

Although the result of the first segmentation has been apparently correct, some pixels that belonged to the skin-line have been misclassified as background and some pixels from pectoral muscle still have remained in the segmented image (Figure 5.1) [110]. Thus, another algorithm has been developed in order to avoid misclassifications of pixels by working on edge detection and scale space concepts [111]. A growing contour process has been established based on concepts of attraction and the regularization from the active contours. The new algorithm has contained the following operations:

1. To generate a scale space representation of the image prior to perform edge detection through different scales.

2. To locate an initial seed point lying in the skin-line contour based on a robustly estimation process. A contour growing process was used based on enlarging and adapting a contour with different criteria through the seed point.

3. To smooth the image.

![Fig. 5.1: Representative result of segmentation. Image enhancement by noise removal and segmentation. (a) Original image; (b) Segmented image](image-url)
5.3 DMC Algorithm

5.3.1 GLCM

In order to use GLCM to obtain the statistical descriptors from all the images with Matlab, two fundamental functions have been used:

- “graycomatrix”: computes the GLC matrix from an image. This matrix can be defined by different parameters: number of gray levels, distance, orientation, and others.

- “graycoprops”: describes the texture of the image by obtaining the statistics: contrast, homogeneity, energy and correlation. As the proportion of white and gray inside the breast is not computed by graycoprops, a customized code has been implemented.

The GLCM code developed includes the following five steps:

1. A graycomatrix has been defined to obtain an 8-level gray image (by default) with distance and orientation [0 3] and symmetric value equal to true (symmetric example: counts both 1,2 and 2,1 pairing when calculating the number of times, the value 1 is adjacent to the value 2).

2. Once computed, the image has been re-scaled by limiting its levels of gray values at 3 (Figure 5.2).

Listing 5.2: Matlab Code: GLCM Obtaining

\[
[glcm,SI] = \text{graycomatrix}(I,'Offset',[0 3],'Symmetric',true);
\]

Listing 5.3: Matlab Code: Image Re-scaling

\[
I1 = SI;  \\
I1 (SI == 1) = 1;  \\
I1 (SI == 2) = 5;  \\
I1 (SI == 3) = 5;  \\
I1 (SI == 4) = 5;  \\
I1 (SI == 5) = 5;  \\
I1 (SI == 6) = 5;  \\
I1 (SI == 7) = 8;  \\
I1 = \text{rescale}(I1);  
\]
3. After obtaining this re-scaled image, another graycomatrix has been computed in order to obtain the statistical descriptors of the image with the function graycoprops:

**Listing 5.4: Matlab Code: Statistical Descriptors Obtaining**

```matlab
[glcm1, I2] = graycomatrix(I1, 'Offset', [0 3], 'Symmetric', true);
stats(count) = graycoprops(glcm, {'contrast', 'homogeneity', 'energy', 'correlation'});
statsTable = struct2table(stats);
White (count) = length(find(I2 == 8));
Gray(count) = length(find(I2 == 5));
Prop(count) = (White/(White+Gray))*100;
```

4. Data distribution has been evaluated by boxploting the results (Figure 5.3).
After analyzing these results, the proportion of white and gray statistic value has revealed the stronger segregation in the different BI-RADS groups. Thus, it has been selected to classify images.
5. Ultimately, classification is done by selecting two values for each BI-RADS category from the proportion of white and gray.

Data description for each BI-RADS category (boxplot) is (Figure 5.4):

![Boxplots of proportions for different BI-RADS categories](image)

**Fig. 5.4**: Matlab Data tips of proportion in images. (a) BI-RADS I; (b) BI-RADS II; (c) BI-RADS III; (d) BI-RADS IV

Hence, the two selected values for each BI-RADS group are (Table 5.3):

<table>
<thead>
<tr>
<th>BI-RADS I</th>
<th>BI-RADS II</th>
<th>BI-RADS III</th>
<th>BI-RADS IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min. proportion</td>
<td>8</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Max. proportion</td>
<td>10</td>
<td>8</td>
<td>15</td>
</tr>
</tbody>
</table>

**Tab. 5.3**: Established proportion limits for each BI-RADS group

**Listing 5.5**: Matlab Code: Classification from Proportion Limits

```matlab
% Defined previously: B1=0;B2=0;B3=0;B4=0;

if (8<Prop#(count))&(Prop#(count)<10)
    B1=B1+1;
elseif (1<Prop#(count))&(Prop#(count)<8)
    B2=B2+1;
elseif (10<Prop#(count))&(Prop#(count)<15)
    B3=B3+1;
elseif Prop#(count)>15
    B4=B4+1;
end
```
5.3. DMC Algorithm

In Prop#, the symbol # refers to the different class of BI-RADS that is evaluated. It can be 1, 2, 3 or 4 depending on the BI-RADS group that is computed. The images have been then classified into the 4 different BI-RADS, and this classification performance has been evaluated through confusion matrix and kappa coefficient in the following sections (Figure 5.5).

![Confusion Matrix](image)

**Fig. 5.5:** GLCM classification. Results of each BI-RADS group to analyze with the confusion matrix

5.3.2 Laws’ Masks

To obtain different texture expression of images, “law” function has been the key part of the code. The function first defines the filters explained in the theoretical part (Subsection 3.4.2) and then these filters are applied to the input images:

```matlab
%% define filters
filters ={};
filters{1}=[1 4 6 4 1];
filters{2}=[-1 -2 0 2 1];
filters{3}=[-1 0 2 0 -1];
filters{4}=[1 -4 6 -4 1];
filters{5}=[-1 2 0 -2 -1];

%% define filters and apply to images
filtered2D ={};
for i=1:size(filters,2)
    for j=1:size(filters,2)
        temp=filters{i}’*filters{j};
        filtered2D{end+1}=imfilter(image,temp);
    end
end
```
Then, 9 different laws’ masks have been obtained from the images:

**Listing 5.7: Matlab Code: Laws’ Masks Obtaining**

```matlab
[mapz] = laws(I);
figure;
subplot(3,3,1); imshow(mapz{1}); axis equal; colormap(gray); colorbar;
subplot(3,3,2); imshow(mapz{2}); axis equal; colormap(gray); colorbar;
subplot(3,3,3); imshow(mapz{3}); axis equal; colormap(gray); colorbar;
subplot(3,3,4); imshow(mapz{4}); axis equal; colormap(gray); colorbar;
subplot(3,3,5); imshow(mapz{5}); axis equal; colormap(gray); colorbar;
subplot(3,3,6); imshow(mapz{6}); axis equal; colormap(gray); colorbar;
subplot(3,3,7); imshow(mapz{7}); axis equal; colormap(gray); colorbar;
subplot(3,3,8); imshow(mapz{8}); axis equal; colormap(gray); colorbar;
subplot(3,3,9); imshow(mapz{9}); axis equal; colormap(gray); colorbar;
suptitle("Laws’ masks", 'FontSize',18);
```

Ultimately, after applying these filters to the image, nine outputs have been obtained and would be the basis for the textural analysis (Figure 5.6).

![Laws’ masks](image)

**Fig. 5.6:** Laws’ masks result. After applying Laws’ filter to an image, this 9 expression are obtained

The next piece of code should design a statistic window computation. Then, statistical descriptors could be evaluated: mean, mean of absolute values, and standard deviation.
Due to time restriction and several unresolved challenges, statistical descriptors have not been obtained.

### 5.3.3 LBP

LBP features have been extracted from the images as indicated in the code below. To process this part of the code, 8 functions have been necessary. Results of computing these functions are shown in Figure 5.7 and Figure 5.8.

**Listing 5.8: Matlab Code: LBP Histograms Obtaining**

```matlab
nFiltSize = 8;

filtR = generateRadialFilterLBP(nFiltSize, nFiltRadius);

effLBP = efficientLBP(I1, 'filtR', filtR, 'isRotInv', false, 'isChanWiseRot', false);
effRILBP = efficientLBP(I1, 'filtR', filtR, 'isRotInv', true, 'isChanWiseRot', false);
uniqueRotInvLBP = findUniqValsRILBP(nFiltSize);
tightValsRILBP = 1:length(uniqueRotInvLBP);
effTightRILBP = tightHistImg(effRILBP, 'inMap', uniqueRotInvLBP, 'outMap', tightValsRILBP);
binsRange = (1:2^nFiltSize) - 1;
```

![Regular LBP histogram](image1)

![RI-LBP sparse histogram](image2)

![RI-LBP tight histogram](image3)

**Fig. 5.7: LBP histograms result**
Listing 5.9: Matlab Code: LBP Images Obtaining

```matlab
tic;

effLBP = efficientLBP(I1, 'filtR', filtR, 'isRotInv', true, 'isChanWiseRot', false);

effTime = toc;

tic;

pwLBP = pixelwiseLBP(I1, 'filtR', filtR, 'isRotInv', true, 'isChanWiseRot', false);

inEffTime = toc;
```

Result obtained after computing this piece of the above code is showed in the following three group of Figure 5.8.

![Original image](image1.png) ![Efficient LBP image](image2.png) ![Pixel-wise LBP image](image3.png)

**Fig. 5.8:** LBP Computation. First shows the original image without being pre-processed. Secondly it is shown the efficient LBP image and thirdly, the pixel-wise LBP image result.

As it was the case of Laws’ mask, unresolved challenges made it impossible to extract statistical descriptors as expected. In order to follow with the analysis of these statistical descriptors, it should also be designed a statistic window computation that would compare different textures detected in breast.
5.4 AlexNet Algorithm

AlexNet algorithm [98] used in this project includes the following analysis steps (Figure 5.9):

![Fig. 5.9: AlexNet architecture performed in this project](image)

5.5 Analysis of the Results: Confusion Matrix

This section covers the methods used for the results analysis in this project: confusion matrices and kappa coefficient. Confusion matrices have been chosen to evaluate the two breast composition classification systems. A confusion matrix is a predictive and useful analytic tool that summarizes a classifier performance. This matrix reveals the number of well and miss classified instances per class and contains information about actual and predicted values in a classification system [77]. Confusion matrix is a widely used tool because it gives a more complex and objective point of view about a classification model.
In a confusion matrix, rows represent instances in a predicted class and columns are instances in the actual class (or vice versa). The four outcomes in a confusion matrix (2x2) are:

- **True positive**: correctly predicted values as positive.
- **True negative**: correctly predicted values as negative.
- **False positive**: incorrectly predicted values. They were originally negative but predicted as positive.
- **False negative**: incorrectly predicted values. They were originally positive but predicted as negative.

![Confusion Matrix Diagram]

**Fig. 5.10**: Confusion matrix. Conditions and different parameters that can be obtained through a confusion matrix

The rates analyzed in a confusion matrix (Figure 5.10) and included in this project are: accuracy, true positive rate, false negative rate, false positive rate, true negative rate, false omission rate, positive predictive value and negative predictive value.

### 5.5.1 Kappa (k) Coefficient for Confusion Matrices

In 1960, Cohen introduced for the first time the kappa (k) coefficient [113], a statistical measure that can be extracted from a confusion matrix to estimate the agreement in categorical data. The common interpretation of k values are shown in Table 5.4, where P(D) is the relative agreement observed among raters (accuracy) and P(E) is the hypothetical probability of chance agreement (Table 5.5).
This interpretation categorizes the agreement in: poor, slight, fair, moderate, substantial and almost perfect. A model is considered statistically perfect when the k coefficient is equal to 1.

**Tab. 5.4:** Coefficient Kappa (k). Common interpretation of k values

<table>
<thead>
<tr>
<th>k</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0</td>
<td>Poor</td>
</tr>
<tr>
<td>[0, 0.20]</td>
<td>Slight</td>
</tr>
<tr>
<td>[0.21, 0.40]</td>
<td>Fair</td>
</tr>
<tr>
<td>[0.41, 0.60]</td>
<td>Moderate</td>
</tr>
<tr>
<td>[0.61, 0.80]</td>
<td>Substantial</td>
</tr>
<tr>
<td>[0.81, 1]</td>
<td>Almost perfect</td>
</tr>
</tbody>
</table>

- P(D): the sum of diagonal terms divided by the sum of total instances (percentage of correctly classified instances).
- P(E): is the probability of random agreement. P(E) is calculated as below:

\[
P(\text{correct}) = \frac{A+B}{A+B+C+D} \times \frac{A+C}{A+B+C+D}
\]

\[
P(\text{incorrect}) = \frac{C+D}{A+B+C+D} \times \frac{B+D}{A+B+C+D}
\]

\[
P(E) = P(\text{correct}) + P(\text{incorrect})
\]

**Tab. 5.5:** Coefficient Kappa simplified example

<table>
<thead>
<tr>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>A</td>
</tr>
<tr>
<td>Incorrect</td>
<td>C</td>
</tr>
</tbody>
</table>
Chapter 6

Results

6.1 Evaluation of Pre-Processed Mammography Images

The mammography images have been pre-processed and filtered-segmented, and then classified into 4 BI-RADS categories. Representative segmented images for these categories are shown in Figure 6.1, Figure 6.2, Figure 6.3 and Figure 6.4.

Figures below are ordered by BI-RADS group and show: (a) Original image; (b) Mask; (c) Segmented image.

Fig. 6.1: Representative filtered-segmented BI-RADS I image
Chapter 6. Results

Fig. 6.2: Representative filtered-segmented BI-RADS II image

Fig. 6.3: Representative filtered-segmented BI-RADS III image

Fig. 6.4: Representative filtered-segmented BI-RADS IV image
6.2 Evaluation of Breast Composition Classification

GLCM and AlexNet confusion matrices and kappa coefficients have been obtained and are shown in Figure 6.5, Figure 6.6 and Figure 6.7.

![Confusion matrix](image)

**Fig. 6.5:** Confusion matrix obtained when using GLCM to classify

**Listing 6.1:** Matlab Code: GLCM Confusion Matrix and Coefficient Kappa Obtaining

```matlab
% GLCM confusion matrix
x0 = [126 60 7 0; 69 131 0 0; 0 0 179 21; 0 0 14 186];

GLCMCM = confusionchart(g2);

kappa(x0);
```

**UNWEIGHTED COHEN’S KAPPA**

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed agreement (po)</td>
<td>0.7844</td>
</tr>
<tr>
<td>Random agreement (pe)</td>
<td>0.2500</td>
</tr>
<tr>
<td>Agreement due to true concordance (po-pe)</td>
<td>0.5343</td>
</tr>
<tr>
<td>Residual not random agreement (1-pe)</td>
<td>0.7500</td>
</tr>
<tr>
<td>Cohen’s kappa</td>
<td>0.7125</td>
</tr>
<tr>
<td>Kappa error</td>
<td>0.0195</td>
</tr>
</tbody>
</table>
kappa C.I. (alpha = 0.0500) = 0.6743 0.7506
Maximum possible kappa, given the observed marginal frequencies = 0.9849
k observed as proportion of maximum possible = 0.7234
Substantial agreement
Variance = 0.0004 z (k/sqrt(var)) = 34.7343 p = 0.0000
Reject null hypothesis: observed agreement is not accidental

As AlexNet algorithm have been performed in two steps, train and test, so two confusion matrices have been obtained. The train subset uses a set of examples to fit the parameters of the model in order to train the network, while the test one is employed to evaluate the network performance on unseen data. As shown in (Figure 6.6), train subset has included 693 mammography images of 4 different classes. Correct classification has been achieved in 87.16% of images.

**Listing 6.2:** Matlab Code: AlexNet train Coefficient Kappa Obtaining

```matlab
% Test confusion matrix
x1=[164 2 5 1;28 138 11 3;2 4 161 10; 3 4 16 1 4 1 ];
kappa(x1)
```

**Fig. 6.6:** Confusion matrix obtained when using AlexNet train algorithm to classify
6.2. Evaluation of Breast Composition Classification

UNWEIGHTED COHEN’S KAPPA

Observed agreement (po) = 0.8716
Random agreement (pe) = 0.2501
Agreement due to true concordance (po-pe) = 0.6215
Residual not random agreement (1-pe) = 0.7499
Cohen’s kappa = 0.8287
kappa error = 0.0169
kappa C.I. (alpha = 0.0500) = 0.7955 0.8620
Maximum possible kappa, given the observed marginal frequencies = 0.9211
k observed as proportion of maximum possible = 0.8997
Perfect agreement
Variance = 0.0005  z (k/sqrt(var)) = 37.4742  p = 0.0000
Reject null hypothesis: observed agreement is not accidental

From the Mamo 1 group (796 images), 693 have been used for training and from the rest, 56 images have been randomly extracted for the test. The 47 remaining images have been discarded due to the poor quality of the images.

Fig. 6.7: Confusion matrix obtained when using AlexNet test algorithm to classify
### Listing 6.3: Matlab Code: AlexNet test Coefficient Kappa Obtaining

```matlab
% Train confusion matrix
x2 = [13 1 0 0; 1 8 5 1; 0 0 6 8; 0 0 5 8];
kappa(x2)
```

**UNWEIGHTED COHEN’S KAPPA**

<table>
<thead>
<tr>
<th>Observed agreement (po) = 0.6250</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random agreement (pe) = 0.2474</td>
</tr>
<tr>
<td>Agreement due to true concordance (po-pe) = 0.3776</td>
</tr>
<tr>
<td>Residual not random agreement (1-pe) = 0.7526</td>
</tr>
<tr>
<td>Cohen’s kappa = 0.5017</td>
</tr>
<tr>
<td>kappa error = 0.0860</td>
</tr>
<tr>
<td>kappa C.I. (alpha = 0.0500) = 0.3332 0.6702</td>
</tr>
<tr>
<td>Maximum possible kappa, given the observed marginal frequencies = 0.8576</td>
</tr>
<tr>
<td>k observed as proportion of maximum possible = 0.5850</td>
</tr>
<tr>
<td>Moderate agreement</td>
</tr>
<tr>
<td>Variance = 0.0063 z (k/sqrt(var)) = 6.3140 p = 0.0000</td>
</tr>
<tr>
<td>Reject null hypothesis: observed agreement is not accidental</td>
</tr>
</tbody>
</table>

### 6.3 Summary of Classification Results

After image enhancement, noise removal, segmentation, statistical values extraction and training the images, a 78.44% and 62.5% of correct classification (accuracy) has been obtained (Table 6.1), which indicates the success of breast tissue correct classification through different methods.

**Table 6.1:** Summary of the classification results. Accuracy obtained and images used with each method

<table>
<thead>
<tr>
<th>GLCM</th>
<th>AlexNet (train)</th>
<th>AlexNet (test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>795 images</td>
<td>78.44%</td>
<td></td>
</tr>
<tr>
<td>693 images</td>
<td>82.16%</td>
<td></td>
</tr>
<tr>
<td>56 images</td>
<td></td>
<td>62.5%</td>
</tr>
</tbody>
</table>
Discussion

The main objective of this project was to classify images by using two algorithms: Digital Mammography Challenge (DMC) algorithm and Convolutional Neural Network (CNN) algorithm after hypothesizing that a CNN classifier would present better results than a traditional classifier (DMC).

Different steps (and challenges) have been taken in order to test this hypothesis. First, images from MIAS database needed to be converted to PNG format, which led to a loss of resolution (reason why they were removed from the analyses). Moreover, a worsen outcome than expected was obtained due to an adaptation of the pre-process code to PNG images (from DICOM images).

Another challenge that I encountered was the fact that although all the images were supposedly categorized correctly after noise reduction and segmentation, this was not the case. That is, some images revealed a mismatch in some pixels from the pectoral muscle considering them as breast tissue and vice versa (Figure X), leading to an incorrect classification. The appearance of muscle tissue provided information that interfered with the ROI information. Furthermore, images that presented a big quantity of dense tissue were more difficult to filter and segment. The higher the density of the breast the lower the contrast among different tissues and structures, ultimately leading to less extraction of information (Figure 7.1).

Due to an incorrect segmentation during pre-processing, the image “A_0477_1.RIGHT-MLO_b1” (from Mamo 1 group) showed an all-black image that altered the feature extraction process. This image was therefore removed from the analyses and not included in the following steps.
To evaluate the results obtained with the GLCM, it was necessary to focus on the statistical descriptors extracted. As seen in section 5.3, correlation, contrast, homogeneity and entropy did not reveal many differences among the BI-RADS categories. This impeded the correct processing of these statistical values and led to use only the proportion of white (dense tissue) in the breast, thus limiting image features treatment and results. Despite this limitation, the proportion of dense tissue in the breast showed significant differences among BI-RADS categories and allowed for a correct classification of the images. However, this classification model has been customized to these images and it was not as general as AlexNet algorithm. Due computer limitations, lack of expertise and time restrictions, Laws’ and LBP methods have been tested but not computed as a feature extractor for classification, reason why they have not been analysed as GLCM in this study.

Regarding the analysis of the results of the classifiers, it is worth mentioning that both algorithms misclassified images of adjacent classes. However, in the test confusion matrix generated from AlexNet, more than the 57.1% from BI-RADS 3 images were misclassified as BI-RADS 4, meaning 8 out of 14 showed an incorrect output. The percentage of accuracy in correct image classification revealed that AlexNet train had a higher number of correctly categorized images (82.16%) in comparison to GLCM and AlexNet test (78.44% and 62.5%, respectively). Of note, the classification accuracy obtained with the AlexNet train algorithm (82.16%) was higher than the AlexNet test (62.5%), as expected. All these results are related to the coefficient kappa that has been calculated in order to evaluate the degree of agreement among raters (Table 7.1).
Tab. 7.1: Summary of the agreement among raters

<table>
<thead>
<tr>
<th></th>
<th>GLCM</th>
<th>AlexNet (train)</th>
<th>AlexNet (test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient Kappa</td>
<td>0.7125</td>
<td>0.8287</td>
<td>0.5017</td>
</tr>
<tr>
<td>Agreement</td>
<td>Substantial</td>
<td>Perfect</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

Altogether, the results presented in this project are in line with the number of correctly categorized images by Christian Mata [83]: 83% and 78% of accuracy with 2 different databases using GLCM texture descriptor (combined with Laws’ and LBP). However, the table of other projects results detailed in section 3.1. showed accuracies between 93-100% [94] (that are far from my results).

In light of these results, the hypothesis is not supported. However, in order to test this hypothesis, an important point should be considered. Finding a method to classify images without artificial intelligence (AI) has been challenging and difficult as all the CAD systems found in other studies used AI. Consequently, DMC has been customized to the given images, leading to better results. Thus, regardless of the main hypothesis (comparing AlexNet and DMC) working on classification of mammography image requires AI.
First, the appearance of pectoral muscle in the incorrectly segmented images had a direct negative effect on the classification results. Therefore, I conclude that a more precise segmentation would provide significantly better results.

Second, GLCM and AlexNet train and test algorithms performed well in image classification considering the size of the dataset and that both GLCM and AlexNet test confusion matrix misclassified images of adjacent BI-RADS classes. Hence, I conclude that a larger dataset would allow AlexNet training algorithm to better detect and generalize problems in images, thus achieving significantly improved results in both performances (train and test).

Ultimately, after studying published literature in depth and related it to the results obtained, I can conclude that Neural Networks analysis is the key tool to solve the controversy generated about effectiveness of mammography screening.
Reflections

I would like to express some of the reflections I made during these months.

First, I would like to emphasize how grateful I am for all the knowledge I have gained during this project, although the initially planned project was changed and adapted. I have become aware of the importance of computer systems that process mammography images daily, representing an important help for radiologists. Hence, during the writing of this project, I have noticed how automatic classification of mammography images is in fact an unresolved problem, reason why I think that research should aim to improve the study of mammography image processing and classification in the future.

After reviewing published literature in depth, I realized how challenging it was to find quality information in scientific articles with open access. Moreover, I have also noted that most of the research papers have been published by men. Indeed, only 28 of the 100 scientific articles had a woman as main author. This confirms the need to fight for more access, presence and opportunities for women in STEM.

Lastly, I would like to emphasize how challenging (both professionally and personally) it has been to complete my project in Turin during the SARS-CoV-2/Covid-19 situation, which basically forced me to change all the planned project.
Bibliography


[31] “USF Digital Mammography Home Page.”


[99] “AlexNet - ImageNet Classification with Convolutional Neural Networks.”

[100] “Understanding AlexNet | Learn OpenCV.”


[102] “A Beginner’s Guide to Understanding Convolutional Neural Networks (Part 2) - DZone AI.”

[103] “Feature Maps. Feature Map is also called as . . . | by Chris Kevin | Medium.”


