

Université MUSTAPHA Stambouli  
Mascara



جامعة مصطفى اسطمبولي  
معسكر

Faculty of Exact Science  
Department of Computer Science  
Laboratory of LYSIS

## DOCTORAL LMD THESIS

Features extraction and medical images  
description for breast cancer automatic diagnosis

**Author: Mohammed EL Amine YERMES**

**Supervisor: Dr. Mohammed DEBAKLA**

27/05/2025

In front of the jury:

President	Mohammed SALEM	Pr	University of MASCARA
Examiner	Boudjellal MEFTAH	Pr	University of MASCARA
Examiner	Kada BENYAHIA	MCA	University of SAIDA
Supervisor	Mohammed DEBAKLA	MCA	University of MASCARA
Co- Supervisor	Khalifa DJEMAL	Pr	University of Paris Saclay Evry, France

2024-2025

*To my parents, nothing can replace your love and sacrifices*

*To my wife and my son Zakaria, you are my world*

*To my brothers and sisters, for your constant support and inspiration*

*To my grandmother and my uncle Lahcène, you mean a lot to me*

*To my family in law, thanks for being part of this*

# Acknowledgements

*First and foremost, Praise to Almighty God for granting me the strength, patience, and perseverance to complete this work. Without His guidance and countless blessings, none of this would have been possible.*

*I would like to extend my sincere and heartfelt thanks to my supervisor, Mr. DEBAKLA Mohammed, for his invaluable support, insightful direction, and continuous encouragement throughout the course of this research. Your expertise and commitment were instrumental to the success of this work.*

*Special thanks also go to my co-supervisor, Mr. DJEMAL Khalifa, whose assistance, advice, and constructive feedback have significantly contributed to the development of this thesis.*

*I am truly grateful to the members of the jury:*

- *Mr. SALEM Mohammed professor at Mustapha Stambouli University of Mascara, Algeria,*
- *Mr. MEFTAH Boudjellal professor at Mustapha Stambouli University of Mascara, Algeria,*
- *Mr. BENYAHIA Kada MCA at Tahar Moulay University of Saïda, Algeria,*

*My appreciation also goes to all the staff of the Computer Science Department. Your help, whether academic, administrative, or technical, has been greatly appreciated.*

*Finally, I would like to thank everyone who contributed to this work in any way, colleagues, friends, and family, your support, encouragement, and kindness have meant a great deal to me.*

*Thank you all.*

# Abstract

## Features extraction and medical images description for breast cancer automatic diagnosis

Fighting breast cancer remains a major public health concern worldwide, affecting millions of lives each year. Early detection is essential to improve survival rates and treatment results. In recent years, advances in medical imaging, particularly mammography, combined with features-based methods, and deep learning techniques, have significantly improved the accuracy and efficiency of computer-aided diagnosis (CAD) systems. Breast masses and microcalcifications represent the most frequent anomalies with a high risk of malignancy. Most of descriptors found in the literature extract global features and fail to characterize spiculated masses.

To address this problem, we focused on developing descriptor adapted to the context of breast cancer, and particularly spiculated masses. PATAR descriptor (Polygon Approximation Triangle-Area Representation) applies a geometric transformation on masses, to simplify the contour while keeping important characteristics like concave and convex spaces. Polygon approximation is done with the Ramer-Douglas-Peucker (RDP) algorithm. After RDP process Triangle-Area Representation (TAR signature) is calculated to quantify and measure spiculations. TAR signature calculates the area made by the corners of polygon.

In recent years, deep learning-based models have gained ground in CADx systems. Models like DenseNet, ResNet, and EfficientNet based on Convolutional Neuronal Networks (CNNs) does not perform well facing microcalcifications. Stacking Ensemble learning is a technique that combine multiple model outputs, through meta-learner to make final prediction. We designed an optimal meta-learner composed of fully connected network. Experiment on CBIS-DDSM dataset demonstrate the efficiency of the meta-learner. Boosting is another ensemble learning strategy that learns multiple models sequentially and adjust samples weights after each iteration. In this context, a new boosting algorithm is proposed named Cost-Sensitive Boosting with Error Weighted Adjustments (CSB-EWA). The main contribution in this algorithm consist in using false positive and false negative rates to adjust samples weight to guarantee maximum balance between sensitivity and specificity.

**Keywords:** Breast cancer, masses, Microcalcifications, Computer-Aided Diagnosis, Features Extraction, Ensemble Deep Learning, Meta-learner, Boosting learning

## ملخص

### استخراج المميزات ووصف الصور الطبية للتشخيص التلقائي لسرطان الثدي

لا تزال مكافحة سرطان الثدي مشكلة صحية عامة كبيرة في جميع أنحاء العالم، حيث تؤثر على حياة الملايين من النساء كل عام. ويُعد الكشف المبكر ضروريًا لتحسين نتائج العلاج ومعدلات البقاء على قيد الحياة. في السنوات الأخيرة، أدى التقدم في مجال التصوير الطبي، لا سيما التصوير الشعاعي للثدي، بالإضافة إلى الأساليب القائمة على المميزات وتقنيات التعلم العميق، إلى تحسين دقة وكفاءة أنظمة التشخيص بمساعدة الحاسوب بشكل كبير. تمثل كتل الثدي وتكلسات الثدي الدقيقة أكثر الحالات الشاذة شيوعًا التي تنطوي على مخاطر عالية للإصابة بالأورام الخبيثة. معظم الواصفات الموجودة في الأدبيات تستخرج السمات العامة وتفشل في توصيف الكتل المتشعبة.

ولمعالجة هذه المشكلة، ركزنا على تطوير واصف يتكيف مع سياق سرطان الثدي، وخاصة الكتل المتشعبة. يطبق الواصف (تقريب المضلع التقريبي للمثلث-المساحة) تحويلًا هندسيًا على الكتل، لتبسيط الشكل الكتل مع الحفاظ على الخصائص المهمة مثل المساحات المقعرة والمحدبة. يتم تقريب المضلع باستخدام خوارزمية Ramer-Douglas-Peucker (RDP) بعد عملية RDP يتم حساب مساحة منطقة المثلث بواسطة توقيع TAR لتحديد وقياس المساحات المقعرة. يحسب توقيع TAR المساحة التي تصنعها زوايا المضلع. المميزات التي تم استخراجها عن طريق هذا الواصف تسمح بتقدير مدى تشعب الكتل وبالتالي تقدير مدى خبث هذه الأورام.

في السنوات الأخيرة، اكتسبت النماذج القائمة على التعلّم العميق مكانة في أنظمة التشخيص التلقائي. لا تؤدي نماذج مثل DenseNet و ResNet و EfficientNet القائمة على الشبكات العصبية التلافيفية (CNNs) أداءً جيدًا في مواجهة التكلسات الدقيقة. التعلّم التجميعي هو تقنية تجمع بين مخرجات نماذج متعددة، من خلال متعلم تلوي للتنبؤ النهائي. لقد صممنا متعلمًا فائقًا مثاليًا يتألف من شبكة متصلة بالكامل. أثبتت التجربة على مجموعة بيانات CBIS-DDSM كفاءة المتعلم الفوقوي. التعزيز هو استراتيجية أخرى للتعلم التجميعي تقوم على تعدد النماذج بالتتابع وتعديل العينات الأوزان بعد كل تكرار. في هذا السياق، اقترحت خوارزمية تعزيز جديدة تسمى التعزيز الحساس للتكلفة مع تعديلات مرجحة للخطأ (CSB-EWA). تتمثل المساهمة الرئيسية في هذه الخوارزمية في استخدام معدلات إيجابية كاذبة ومعدلات سلبية كاذبة لضبط وزن العينات لضمان أقصى قدر من التوازن بين الحساسية والخصوصية.

**الكلمات المفتاحية:** سرطان الثدي، الكتلة، التكلسات الدقيقة، التشخيص بمساعدة الحاسوب، استخراج المميزات، مجموعة التعلم العميق، المتعلم الفوقوي، التعزيز

## Résumé

### Extraction de caractéristiques et description d'images médicales pour le diagnostic automatique du cancer du sein

La lutte contre le cancer du sein reste un problème majeur de santé publique dans le monde entier, affectant des millions de vies chaque année. Le dépistage précoce est essentiel pour augmenter les chances de succès du traitement. Ces dernières années, les progrès de l'imagerie médicale, en particulier la mammographie, combinés aux méthodes basées sur les caractéristiques et aux techniques d'apprentissage profond, ont considérablement amélioré la précision et l'efficacité des systèmes de diagnostic assisté par ordinateur (DAO). Les masses mammaires et les microcalcifications représentent les anomalies les plus fréquentes avec un risque élevé de malignité. La plupart des descripteurs trouvés dans la littérature extraient des caractéristiques globales et ne parviennent pas à caractériser les masses spiculées.

Pour résoudre ce problème, nous nous sommes concentrés à développer un descripteur adapté au contexte du cancer du sein, et en particulier aux masses spiculées. Le descripteur PATAR (Polygon Approximation Triangle-Area Representation) applique une transformation géométrique aux masses, afin de simplifier le contour tout en conservant les caractéristiques importantes telles que les espaces concaves et convexes. L'approximation en polygone est effectuée à l'aide de l'algorithme Ramer-Douglas-Peucker (RDP). Après le processus RDP, la signature TAR est calculée pour quantifier et mesurer les spiculations. La signature TAR calcule la surface formée par les coins d'un polygone ce qui permet d'estimer le degré de malignité.

Ces dernières années, les modèles basés sur l'apprentissage profond ont gagné du terrain dans les systèmes CADx. Les modèles tels que DenseNet, ResNet et EfficientNet, basés sur des réseaux de neurones convolutifs (CNN), ne donnent pas de bons résultats face aux microcalcifications. L'apprentissage d'ensemble par empilement est une technique qui combine les résultats de plusieurs modèles, par le biais d'un meta-learner, afin d'obtenir une prédiction finale. Nous avons conçu un meta-learner optimal composé des couches entièrement connecté. Des expériences sur la base de données CBIS-DDSM démontrent l'efficacité du méta-apprentissage. Le boosting est une autre stratégie d'apprentissage d'ensemble qui permet d'apprendre plusieurs modèles de manière séquentielle et d'ajuster les poids des échantillons après chaque itération. Dans ce contexte, un nouvel algorithme de boosting est proposé à l'adresse sous le nom de Cost-Sensitive Boosting with Error Weighted Adjustments (CSB-EWA). La principale contribution de cet algorithme consiste à utiliser les taux de faux positifs et de faux négatifs pour ajuster le poids des images afin de garantir un équilibre maximal entre la sensibilité et la spécificité.

**Mots-clés :** Cancer du sein, Masses, Microcalcifications, Diagnostic assisté par ordinateur, Extraction d'attributs, apprentissage profond d'ensemble, meta-learner.

# List of Figures

<b>Figure 1.1</b> : Breast anatomy. ....	7
<b>Figure 1.2</b> : The different types of microcalcifications: a) coarse or coralliform, b) cutaneous or dermal, c) vascular, d) round, e) clear center, f) rod like, g) eggshell or parietal, h) milk-calcium-like, i) suture-like, j) dystrophic, k) amorphous or indistinct, l) polymorphic fine, m) branched linear. ....	9
<b>Figure 1.3</b> : The various distributions of microcalcifications: a) Segmental, b) Diffuse or scattered, c) Clustered, d) Regional, and e) Linear. ....	11
<b>Figure 1.4</b> : The four shapes of a breast mass: a) Round, b) Oval, c) Lobulated, and d) Irregular. ....	12
<b>Figure 1.5</b> : Different contour shapes for a mass: a) Circumscribed, b) Microlobulated, c) Masked, d) Indistinct, and e) Spiculated ....	13
<b>Figure 1.6</b> : BIRADS classification of breast density: a) Density < 25%, b) Density between 25 and 50%, c) Density between 50 and 75% and d) Density > 75%. ....	15
<b>Figure 1.7</b> : Example of an architectural distortion. ....	15
<b>Figure 1.8</b> : Left and right profile mammogram showing asymmetry of shape. ....	16
<b>Figure 1.9</b> : Examples of breast ultrasound: a) ultrasound with abnormality (mass), b) normal breast. ....	17
<b>Figure 1.10</b> : Example of breast MRI bi-sein. ....	18
<b>Figure 1.11</b> : Thermography image. ....	19
<b>Figure 1.12</b> : Positron Emission Tomography image. ....	20
<b>Figure 1.13</b> : An example of histopathological image. ....	21
<b>Figure 1.14</b> : The different mammography views, with MLO and CC being the most commonly used. ....	22
<b>Figure 2.1</b> :Workflow of Computer-Aided Diagnosis and Detection systems [178]. ....	27
<b>Figure 3.1</b> : Rectangularity measure. ....	48
<b>Figure 3.2</b> : PATAR descriptor process, starting with polygon approximation applied on mass ROI's then TAR signature to extract features generation. ....	63
<b>Figure 3.3</b> : Example of Polygon approximation: a) the original mammographic mass, b) CBIS-DDSM ROI, c) contour and shape of the mass and d) polygon approximation of the mass. ....	64

<b>Figure 3.4:</b> Different $\epsilon$ values applied to the same mass contour: (a) the original mass contour without approximation; (b) with $\epsilon = 0.005$ ; (c) with $\epsilon = 0.010$ ; and (d) with $\epsilon = 0.015$ .....	66
<b>Figure 3.5:</b> Three possible outcomes for the TAR signature are observed: the contour is convex when TAR is positive, concave when TAR is negative, and forms a straight line when TAR equals zero.....	67
<b>Figure 3.6:</b> Number of corners as feature explain clear variance between (a) Benign mass with round shape having 13 corners, and (b) malignant mass with irregular shape having 25 corners. ....	68
<b>Figure 3.7:</b> Illustration of the distinction between a round mass (a) and a spiculated mass (b) based on the values of TARN (green triangles) and TARP (black triangles).....	69
<b>Figure 3.8:</b> The proposed approach builds a feature model using training data, applies SVM, FCM, and RF classifiers to classify malignant vs. benign masses, and evaluates PATAR's diagnostic performance during testing. ....	72
<b>Figure 3.9:</b> Four cropped images of mass from CBIS-DDSM dataset, a) mammographic images without segmentation, b) shows segmented mass (Mask image). ....	74
<b>Figure 3.10:</b> Comparing the best results of PATAR, with and without polygon approximation.....	77
<b>Figure 3.11:</b> ROC curve and confusion matrix of RF, SVM and FCM using CBIS-DDSM. Random Forest classifier ensures the best results in terms of accuracy and AUC. ....	79
<b>Figure 3.12:</b> Evolution of PATAR performance cross different values of Epsilon, all metrics decreased when epsilon exceeds 0.01 the universal value. ....	80
<b>Figure 4.1:</b> Relation between Artificial Intelligence, Machine Learning and Deep Learning.	87
<b>Figure 4.2:</b> Architecture of Deep neuronal network, composed of input layer, multiple hidden layers and output layer. ....	88
<b>Figure 4.3:</b> Graphic representation of Sigmoid, ReLU and Leaky ReLU.....	91
<b>Figure 4.4:</b> Architecture of a single layer perceptron (SLP).....	93
<b>Figure 4.5:</b> (a) RNNs with feedback connections, (b) FNNs process inputs independently without memory of past data. ....	96
<b>Figure 4.6:</b> LSTM units that forms the basic Long-Short Term Memory Networks architecture .....	97
<b>Figure 4.7:</b> Convolution operation, the core concept of CNNs.....	98
<b>Figure 4.8:</b> General structure of an individual CNN network with $\alpha$ convolution blocks. ....	99
<b>Figure 4.9:</b> Deep learning application in medical imaging [141]. ....	99
<b>Figure 4.10:</b> Bagging ensemble learning approach.....	103

<b>Figure 4.11:</b> Boosting Approach in Ensemble Learning.....	103
<b>Figure 4.12:</b> Architecture of stacking ensemble model using meta learner.....	104
<b>Figure 4.13:</b> Hard and soft voting.....	105
<b>Figure 4.14:</b> Blending strategy used ensemble learning.....	105
<b>Figure 4.15:</b> (a) Suspicious abnormality detected in image (a), after segmentation in image (b), ROI image in image (c) is extracted to be used for characterization by deep learning model.....	109
<b>Figure 4.16:</b> Class distribution in training and test dataset of CBIS-DDSM.....	109
<b>Figure 4.17:</b> Image samples taken from CBIS-DDSM dataset.....	110
<b>Figure 4.18:</b> Different preprocessing techniques applied on mammograms.....	111
<b>Figure 4.19:</b> Concept of Residual blocks in ResNet networks.....	113
<b>Figure 4.20:</b> ResNet-50 architecture.....	113
<b>Figure 4.21:</b> DenseNet-121 architecture.....	114
<b>Figure 4.22:</b> EfficientNet-b0 architecture.....	116
<b>Figure 4.23:</b> Confusion matrix and ROC curve of ResnNet18, DenseNet-121, and EfficientNet-b0.....	117
<b>Figure 4.24:</b> Composition of the proposed stacking system, composed of EfficientNet-b0, DenseNet-121 and ResNet-50 models and optimized meta learner.....	118
<b>Figure 4.25:</b> Architecture of the proposed stacking model with Meta-learner.....	119
<b>Figure 4.26:</b> Confusion matrix and ROC Curve of the Stacking Ensemble model proposed with optimal Meta Learner.....	121
<b>Figure 4.27:</b> Algorithm diagram of Cost-Sensitive Boosting with Error Weighted Adjustments.....	128
<b>Figure 4.28:</b> Boosting ensemble learning system using CSB-EWA algorithm.....	129
<b>Figure 4.29:</b> Results obtained with first scenario: Boosting of the ensemble learning using Adaboost.....	130
<b>Figure 4.30:</b> Results obtained with second scenario: Boosting of the ensemble learning using Gradient Boost.....	131
<b>Figure 4.31:</b> Results obtained with third scenario: Boosting of the ensemble learning using proposed CSB-EWA algorithm.....	131
<b>Figure 4.32:</b> Histogram of metrics obtained with three scenarios and compared with performance of base-models ResNet-50, DenseNet-121 and EfficientNet-b0.....	133

## List of Tables

<b>Table 1.1 :</b> Comparison between different screening modalities. ....	24
<b>Table 1. 2 :</b> Interpretation and action to be taken (CAT) according to BI-RADS classes. ....	24
<b>Table 2.1 :</b> Confusion matrix table. ....	42
<b>Table 3.1:</b> Evaluation of different methods for mass description. ....	62
<b>Table 3.2:</b> Number of Cases, Abnormalities and Images in the Training and Test Sets, each case can have one or more abnormalities and more images.....	73
<b>Table 3.3:</b> Comparison of results obtained with and without polygon approximation. ....	79
<b>Table 3.4:</b> PATAR performance using Random Forest classifier with different values of Epsilon.....	80
<b>Table 3.5:</b> Comparison of Methods for Mammographic Mass Characterization.....	81
<b>Table 3.6:</b> Performance comparison between proposed approach and some previous works.	82
<b>Table 4.1:</b> Principal commercialized AI-based tools for breast cancer [141].....	100
<b>Table 4.2:</b> Performance of base-models .....	117
<b>Table 4.3:</b> Comparison between the proposed stacking system and individual base-models performances. ....	121
<b>Table 4.4:</b> Comparison between proposed model and some previous works. ....	122
<b>Table 4.5:</b> Different characteristics of models used in our experimentation.....	129
<b>Table 4.6:</b> Summary of all metrics and results obtained in experiments process.....	132

## List of Acronyms

ACR	American College of Radiology
AHE	Adaptive histogram equalization
AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC	Area Under Curve
BIRADS	Breast-Imaging Reporting and Data System
BRCA1	BReast CAncer gene 1
BRCA2	BReast CAncer gene 2
CADe	Computer-Aided Detection
CADx	Computer-Aided Diagnosis
CBIS-DDSM	Curated Breast Imaging Subset - Digital Database for Screening Mammography
CC	CranioCaudal
CDR	Cancer Detection Rate
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
CSB-EWA	Cost-Sensitive Boost with Error Weighted Adjustments
CT	Computed Tomography
DBT	Digital Breast Tomosynthesis
DCNN	Deep Convolutional Neural Network
DDSM	Digital Database for Screening Mammography
DFT	Discrete Fourier Transform
DL	Deep Learning
DQN	Deep Q-Network
DWT	Discrete Wavelet Transform
FDA	Food and Drug Administration
FN	False Negative
FNN	Feedforward Neural Network
FP	False Positive
GAN	Generative Adversarial Network
GRU	Gated Recurrent Unit
HE	Histogram Equalization
IDFT	Inverse Discrete Fourier Transform
LLM	Large Language Model
LR	logistic regression
LSTM	Long short-term memory
ML	Machine Learning
MLO	mediolateral oblique
MLP	Multi-Layer Perceptron
MRF	Markov Random Field
MRI	Magnetic Resonance Imaging
NLP	natural language processing

PATAR	Polygon Approximation Triangle-Area Representation
PCA	Principal Component Analysis
PET	Positron Emission Tomography
RBF	Radial Basis Function
RDP	Ramer–Douglas–Peucker
ReLU	Rectified Linear Unit
RF	Random Forest
RL	Reinforcement Learning
RNN	Recurrent Neural Network
ROC	Receiver-Operating characteristic
ROI	Region Of Interest
SLP	Single Layer Perceptron
SMOTE	Synthetic Minority Oversampling Technique
SOM	Self-Organizing Maps
SVM	Support Vector Machine
TAR	Triangle-Area Representation
TN	True Negative
TP	True Positive
VGG	Visual Geometry Group
WT	Wavelet Transform
ZM	Zernike Moments
MCC	Micro Calcification Cluster

# Contents

**Acknowledgements**

**Abstract**

**List of Figures**

**List of Tables**

**List of Acronyms**

**Introduction** ..... 1

## **Chapter 1: Introduction to breast cancer**

1. Introduction	5
2. Breast cancer	5
2.1 Breast anatomy	5
2.2 Breast abnormalities	7
2.2.1 Microcalcifications	7
2.2.2 Masses	12
2.2.3 Architectural distortions	15
2.2.4 Breast asymmetry	15
3. Screening methods	16
3.1 Ultrasound	17
3.2 Magnetic resonance imaging	18
3.3 Tomosynthesis	19
3.4 Thermography	19
3.5 Positron Emission Tomography	20
3.6 Histopathology	20
3.7 Mammography	21
3.7.1 Mammography incidences	21
3.7.2 Screening and diagnostic mammography	23
3.8 Evaluation of different modalities	23
4. The BI-RADS Classification	24
5. Computer aided diagnosis and detection systems	24
6. Conclusion	25

## **Chapter 2: CADx systems for breast cancer**

1. Introduction.....	26
2. CADx systems.....	26
2.1. Definition.....	27
2.1 Data Acquisition.....	28
2.2 Preprocessing.....	28
2.2.1 Noise Reduction.....	29
2.2.2 Contrast Enhancement.....	31
2.2.3 Artifact removal.....	33
2.2.4 Image normalization.....	33
2.2. Segmentation.....	34
2.3.1 Thresholding-based methods.....	34
2.3.2 Region-based segmentation.....	35
2.3.3 Edge-Based Segmentation.....	35
2.3.4 Clustering-based segmentation.....	35
2.3.5 Deep learning-based segmentation.....	36
2.4 Features extraction.....	36
2.4.1 Features.....	36
2.4.2 Feature selection.....	37
2.5 Classification.....	38
2.5.1 Supervised Classification.....	39
2.5.2 Unsupervised Classification.....	39
2.5.3 Reinforcement Learning.....	39
2.5.4 Challenges of classification.....	40
2.6 Metrics and evaluation of CADx systems.....	41
2.6.1 Classification Metrics.....	41
2.6.2 Segmentation Metrics.....	43
2.6.3 Evaluation Strategies.....	43
3. Conclusion.....	44

## **Chapter 3: Image description model for breast cancer recognition**

1. Introduction.....	45
2. Shape descriptors (Morphological).....	46
2.1 Geometric descriptors.....	46
2.1.1 Center of Gravity (Centroid).....	46
2.1.2 Perimeter.....	46

2.1.3 Area.....	46
2.1.4 Circularity .....	47
2.1.5 Rectangularity .....	47
2.1.6 Compactness .....	48
2.1.7 Fourier descriptor .....	48
2.2 High order descriptors .....	49
2.2.1 Moments based descriptors .....	49
2.2.2 Fractal analysis.....	51
2.2.3 Skeletonization-Based Descriptors .....	51
3. Texture descriptors .....	53
3.1 Structural methods.....	53
3.1.2 Morphological Analysis (Mathematical Morphology) .....	53
3.2 Statistical methods.....	54
3.2.1 First-Order Statistical Methods (Histogram-Based Analysis) .....	54
3.3 Model-Based Texture Descriptors .....	55
3.3.1 Markov Random Field (MRF) Models .....	55
3.3.2 Autoregressive (AR) Models .....	55
3.4 Transform Models .....	56
3.4.1 Wavelet Transform (WT).....	56
3.4.2 Curvelet Transform (CT) .....	56
3.4.3 Gabor filters .....	57
4. Evaluation and characteristics of descriptors .....	57
4.1 Characteristics of good descriptors .....	58
4.1.1 Scale Invariance.....	58
4.1.2 Rotation Invariance.....	58
4.1.3 Uniqueness (Distinctiveness).....	58
4.1.4 Robustness to Noise.....	58
4.1.5 Illumination Invariance.....	59
4.1.6 Computational Efficiency .....	59
4.1.7 Discriminative Power .....	59
4.1.8 Comprehensively Descriptive.....	59
4.2 Evaluation.....	59
5. Proposed model for breast mass description (PATAR descriptor) .....	62
5.1 Polygon approximation.....	63
5.2 Triangle-area Representation (TAR) signature .....	66

5.3 Feature extraction .....	68
5.4. Mass classification.....	69
5.4.1 Fuzzy Clustering .....	70
5.4.2 Support Vector Machine .....	71
5.4.3 Random Forest Classifier.....	71
5.5 Experiments and results .....	71
5.5.1 CBIS-DDSM dataset.....	72
5.5.2 Experimentation process .....	74
5.5.2.1 Preprocessing .....	74
5.5.2.2 Features extraction (PATAR Descriptor) .....	74
5.5.2.3 Classification.....	75
5.5.3 Results and discussion .....	75
6. Conclusion.....	83

## **Chapter 4: Automatic diagnosis of microcalcifications using Deep Learning**

1.Introduction.....	84
2. Deep Learning.....	85
2.1. A brief history.....	85
2.1.1 The Birth of AI (1940s – 1950s).....	85
2.1.2 Early AI and the First AI Winter (1957 – 1970s) .....	85
2.1.2 Neural Networks & The Second AI Winter (1980s – 1990s).....	86
2.1.3 Rise of Deep Learning (2000s – 2010s).....	86
2.1.4 Present: AI Revolution (2020s).....	86
2.2 Definitions .....	86
2.2.1 Artificial Intelligence (AI).....	87
2.2.2 Machine Learning (ML) .....	87
2.2.3 Le Deep Learning (DL) .....	87
2.3 Neuronal networks concepts .....	88
2.3.1 Architecture Layers .....	88
2.3.2 Connections .....	89
2.3.3 Activation Functions.....	89
2.3.4 Training.....	91
2.3.5 Prediction.....	92
2.4 Different Types of Neural Networks in Deep Learning .....	92
2.4.1 Perceptron .....	92
2.4.2 Feedforward Neural Networks (FNNs) Network .....	93

2.4.3 Radial Basis Function (RBF) Neural Network .....	94
2.4.4 Recurrent Neural Network (RNNs) .....	95
2.4.5 Long Short-Term Memory (LSTM) Networks.....	96
2.4.6 Convolution Neural Network (CNN) .....	97
2.5 Deep learning in medical imaging and breast cancer .....	99
3. Ensemble Deep Learning.....	102
3.1 Bagging (Bootstrap Aggregating) .....	102
3.2 Boosting.....	103
3.3 Stacking (Stacked Generalization) .....	104
3.4 Voting and Averaging .....	104
3.5 Blending.....	105
3.6 Knowledge Distillation as an Implicit Ensemble Learning Method .....	105
4. Classification of Microcalcifications using Ensemble Learning.....	106
4.1 Related works .....	106
4.2 Data description.....	108
4.3 Pre-processing .....	110
4.3.1 Noise Reduction.....	110
4.3.2 Normalization and Standardization.....	111
4.3.3 Contrast Enhancement .....	111
4.4 Overview of CNN architectures employed in our ensemble learning strategy .....	112
4.4.1 ResNet-50 .....	112
4.4.2 DenseNet-121 .....	113
4.4.3 EfficientNet-b0 .....	114
4.4.4 Experimental evaluation of three Individual CNN models for microcalcifications	116
4.5 Proposed stacked generalization strategy .....	118
4.5.1 Description of stacking model architecture .....	119
4.5.2 Experiments .....	120
4.6 Proposed boosting strategy .....	123
4.6.1 Boosting algorithms in the literature .....	124
4.6.1.1 Adaptive Boosting (AdaBoost) .....	124
4.6.1.2 Gradient Boosting (GB) .....	125
4.6.1.3 Extreme Gradient Boosting (XGBoost) .....	125
4.6.2 Proposed CSB-EWA boosting algorithm.....	125
4.6.3 Experiments and results .....	129
5. Conclusion.....	133

**Conclusion and Perspectives.....134**

# Introduction

Breast cancer represents one of the significant problems in the field of public health. According to World Health Organization breast cancer caused 670 000 deaths globally in 2022 [1]. Indeed, it is the leading cause of cancer-related deaths among females. Breast cancer is a pathological condition characterized by the dysregulated proliferation of abnormal cells within the breast tissue, which can lead to the formation of malignant tumors [2]. If not detected and treated in time, these tumors may invade surrounding tissues and metastasize to other parts of the body, posing a serious threat to life. Manual analysis of medical images is challenging due to the complexity of breast tissue, subtle features like microcalcifications and spiculated masses, and the high risk of human error and variability among radiologists [3]. It is also time-consuming, especially in large-scale screenings. These difficulties underline the importance of Computer Aided Diagnosis (CADx) systems in supporting radiologists in increasing diagnostic accuracy and efficiency. Early detection and diagnosis are critical in improving women's health and reducing mortality.

Mammography, among many technologies such as MRI, Tomography, Ultrasound, or Tomosynthesis, is currently the most widely used screening technique for identifying suspicious lesions such as masses and microcalcifications, which are the most important findings when evaluating breast abnormalities. Mammography, a specialized low-dose X-ray, can detect suspicious abnormalities like masses and microcalcifications deeper in the breast and early stage. However, interpreting mammograms is complex and often subject to human error due to the low contrast of images. In this context, computer-aided diagnosis (CADx) systems have gained increasing attention as a valuable second-reading tool to assist radiologists in accurately detecting and classifying breast abnormalities [4]. Clinical studies demonstrates that CAD systems in breast cancer can improve specificity and sensitivity without a significant increase in the work-up rate, which includes, in most cases, unnecessary biopsies.

Over the years, two major computational approaches have emerged for automatic breast cancer diagnosis: feature-based methods and deep learning-based models. Traditional feature-based methods rely on handcrafted features extracted from images, such as shape, texture, or intensity descriptors. Features are measurable characteristics or attributes extracted from medical images that help describe specific regions of interest (ROI's), such as lesions or tumors, these are then fed into machine learning classifiers. Feature-based methods are more easily interpretable by human experts. This transparency is especially valuable in clinical contexts, where understanding the basis of a diagnosis is essential. These methods typically extract features using global descriptors like Gabor filters, histograms of gradients, or statistical moments. However, such features may not capture the subtle, irregular, or high-level patterns in medical images. Spiculated masses represent a high risk of malignancy when found in the breast [5]. The identification and measurement of these spiculations remain very difficult and challenging due to these masses' hazardous irregularities and random shapes.

Several approaches in the literature failed to characterize efficiently spiculated masses with high accuracy, and the main problem remains in the isolation and quantification of spicules. The design of descriptors specific to the problem of automatic diagnosis of breast cancer is the optimal solution we propose to improve the performance of CADx. In this thesis, our first contribution is developing a robust descriptor named PATAR (Polygon Approximation Triangle-Area Representation). Our descriptor is capable of extracting high-level features such as spiculations and measuring the degree of malignancy in masses. To address this problem, a geometric transformation via polygon approximation is applied on masses as a first step, allowing the descriptor to isolate the most important irregularities and ignore minor variations in contours. The Ramer-Douglas-Peucker (RDP) algorithm is used to approximate the shape of masses to polygon. The RDP algorithm simplifies a curve by retaining only the most significant points. It starts by connecting the first and last points with a straight line, then identifies the point farthest from this line. If the distance exceeds a set tolerance ( $\epsilon$ ), that point is kept, and the process is applied recursively to the sub-curves. If all points lie within the tolerance, the curve is approximated by the straight line. The result is a simplified contour that preserves the shape with fewer points. The RDP algorithm reduces data size and computational cost while preserving essential shape characteristics. Its ability to retain the key structure of a curve makes it valuable in applications such as contour simplification, and medical shape analysis, particularly for simplifying anatomical boundaries like tumors or lesions. Triangle-area representation (TAR signature) is calculated in the second step to quantify the spiculations. TAR signature brows all the points (corners) generated by the RDP algorithm, and calculate the area made by concave and convex spaces. In fact, concavities and convexities are the spicules of original mass. After the polygon approximation and TAR calculation, a vector of features is generated and provided to three classifiers, Support Vector Machines, Random Forest, and Fuzzy C-Means. The high quality of features extracted through PATAR will significantly impacts the accuracy of the proposed machine learning model. Experiments of PATAR are conducted on CBIS-DDSM using 1 545 images.

Deep learning approaches, particularly convolutional neural networks (CNNs), have demonstrated remarkable success in medical imaging [6]. CNNs are especially suitable for image analysis because they preserve spatial relationships through convolutional layers, effectively allowing them to learn hierarchical features, from simple edges to complex patterns. Many model families have been developed, like VGG, ResNet, DenseNet, EfficientNet, and AlexNet. Each model has its strengths and weaknesses. In the automatic diagnosis of microcalcifications, standalone models in the literature struggle to face the low contrast and tiny calcium deposits that form microcalcifications. Most CNNs models use pooling layers that reduce the resolution of the input image, which can cause them to miss small details like microcalcifications [6]. Ensemble learning is a new trend combining multiple models to produce more accurate and robust predictions. Different strategies are introduced to integrate models; stacking is a popular approach involving a meta-learner, classifier, or regressor trained to combine these outputs and learn how to combine the predictions of several base-models [7]. Our second contribution in this thesis comprises the design of an optimal meta-learner, capable of classifying microcalcifications using the predictions of three different models: ResNet-50, DenseNet-121, and EfficientNet-b0. The architecture of the proposed meta-learner is founded

on fully connected layers with ReLU activation functions, and Softmax layer to generate final prediction. CNNs can learn complex correlations between base model outputs permitting the enhancement of the overall performance. The results obtained prove the success of the meta-learner to combine heterogeneous models and integrating different types of outputs.

In Ensemble Learning, boosting strategy consists on training multiple models sequentially, with attribution of new weights to samples after each iteration. The new weights are calculated based on the misclassifications of individual models. Several algorithms are used to generate new weights such as AdaBoost and Gradient Boost, and most of them does not decrease false positive and false negative rate [8]. After testing many models like ResNet-50, DenseNet-121 and EfficientNet-b0, the major problem faced, is the imbalance between sensitivity and specificity in all models. To guarantee a high balance between false positive and false negative rates, boosting strategy is adopted. The third contribution in this work presents new algorithm named Cost-Sensitive Boosting with Error Weighted Adjustments (CSB-EWA). Our solution involves more powerful weights adjustments using  $\lambda_{FP}$  and  $\lambda_{FN}$  penalties calculated using false positive and false negative rates, to make models focus more on misclassified samples. Three scenarios were tested to evaluate the algorithm. First, models have been tested each one individually. Then, boosting ensemble learning using AdaBoost and Gradient Boost are experimented. The third scenario used our proposed algorithm. Experiments used CBIS-DDSM dataset with 1 550 images, and the obtained results confirm the efficiency of the proposition.

Our thesis is organized into introduction, four main chapters, and conclusion.

The introduction presents the general context of the research, the objectives and motivations of the thesis.

Chapter 1 introduces the medical aspects of breast cancer, starting with the anatomy of the breast and abnormalities related to it. Then, a detailed overview of breast masses and microcalcifications is aborded, highlighting their characteristics. Different screening modalities such as ultrasound and mammography will be presented with most important views like Cranio-caudal incidence (CC) and Medio-lateral oblique incidence (MLO). The overview presented in this chapter provides a general comprehension of the context of the problem treated in this thesis.

Chapter 2 is an exhaustive study and detailed presentation of Computer-Aided Diagnosis (CADx) systems and all their components, including segmentation, feature extraction and selection, and classification with careful intention on methods and techniques employed in each step.

Chapter 3's opening sections provide a taxonomy of descriptors commonly discussed in the literature, focusing on shape and geometric descriptors derived and used from fundamental measurements, also texture-based features will be detailed with some descriptors used in medical imaging analysis and particularly breast cancer. Then, the chapter introduces the methodology behind the proposed descriptor PATAR, detailing its mathematical formulation and evaluating its performance compared to existing shape descriptors.

Chapter 4 provides a comprehensive overview of deep learning model architectures, key concepts, and the specific challenges associated with their application in the field of medical

imaging specifically the automatic diagnosis of microcalcifications. It begins by introducing the foundational elements of deep learning, including neural networks, convolutional layers, activation functions, and training strategies. The chapter then explores widely used architectures such as CNNs, RNNs, LSTM, highlighting their design principles and suitability for various medical imaging tasks like classification, segmentation, and detection. Deep learning-based solutions for medical imaging will be discussed and their performances against breast cancer. In this chapter, a novel meta-learner is proposed to combine the predictions of an ensemble learning model. The experiments conducted on CBIS-DDSM show an improvement in the performance of the ensemble model. Also, the proposed algorithm CSB-EWA for boosting ensemble learning is detailed with experiments.

Finally, we conclude our thesis with a general conclusion summarizing our contributions with perspectives and future works.

# Chapter 1: Introduction to breast cancer

## 1. Introduction

Understanding the approach and medical aspects of breast cancer is essential for the optimal design of an automatic diagnostic system. It provides us with the knowledge we need to carry out our study. As one of the most frequent and widely studied cancers, its biological and clinical complexity requires an in-depth understanding to guide tools to help radiologists perfect their diagnosis effectively.

The purpose of this chapter is to provide a comprehensive overview of breast cancer, exploring its generalities and essential features. We begin with the concept of breast cancer, then move on to the anatomical aspect of the disease, as well as associated abnormalities, risk factors, and the various screening and diagnostic methods, with an emphasis on mammography as the most widely used and reliable imaging modality.

This introduction will serve as a starting point for a detailed exploration of the various facets of breast cancer, enabling the reader to better grasp the complexity and importance of this pathology in the context of our research.

## 2. Breast cancer

Breast cancer is a disease marked by the development of malignant (cancerous) cells in breast tissue [8]. These cells undergo an uncontrolled proliferation of cells that can form a mass or tumor. The breast comprises lobules (milk-producing glands), ducts (tubes transporting milk to the nipple), and connective tissue. Most breast cancers start in the cells of the ducts (ductal carcinoma) or lobules (lobular carcinoma), although other types of breast cancer can also develop.

Breast cancer results from a genetic mutation in the breast cells, which a variety of factors can influence. Inherited mutations in the BRCA1 and BRCA2 genes significantly increase the likelihood of breast and ovarian cancer. Having a close relative with the disease can also increase this risk [10]. The risk of developing breast cancer increases with age. Long exposure to estrogen, whether through early menstruation, late menopause, or hormone replacement therapy, can also be a contributing factor. In addition, smoking, excessive alcohol consumption, obesity, and lack of exercise may play a role in increasing risk.

Signs and symptoms of breast cancer may include a lump or thickening in the breast or under the arm, a change in the size, shape, or appearance of the breast, discharge from the nipple other than breast milk, sometimes bloody, redness or orange skin texture on the breast, and unusual pain or tenderness in the breast [11].

### 2.1 Breast anatomy

Breast anatomy plays a crucial role in the context of automated breast cancer diagnosis, as a detailed understanding of its structures is essential for the development and application of

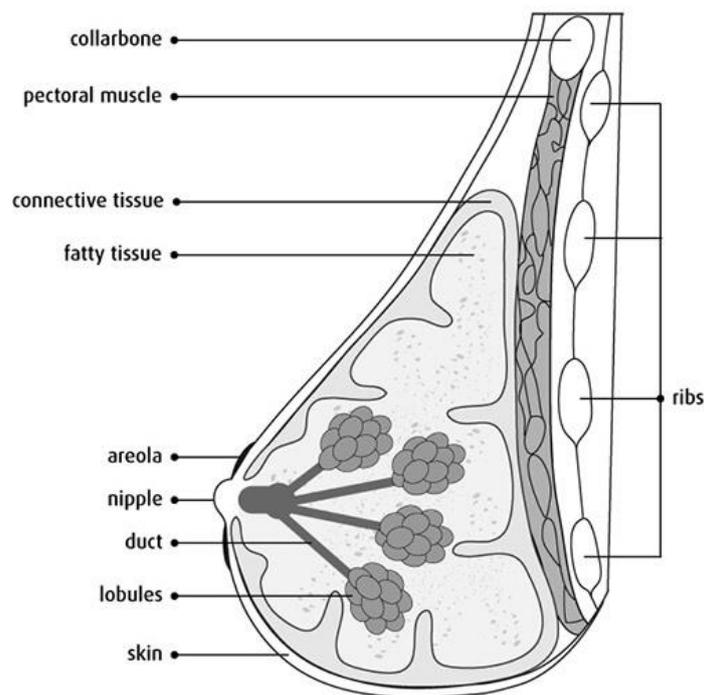
CADx. The breasts, located on the anterior chest wall between the second and sixth ribs, rest on the pectoralis muscle and are surrounded by adipose and connective tissue [12]. This complex structure as shown in Figure 1.1 includes lobules, milk ducts, adipose tissue, connective tissue, blood vessels, lymphatic vessels, and nerves, all of which are critical for detecting and analyzing breast abnormalities using CADx.

The lobules, milk-producing units, are the main sites of the development of lobular carcinoma, a form of breast cancer. Each breast contains 15 to 20 lobes, divided into numerous smaller lobules, with milk-producing cells under hormonal influence [13]. Milk ducts, transporting milk from the lobules to the nipple, are often implicated in ductal carcinoma, the most common form of breast cancer. These ducts branch like tree branches and converge towards the nipple, dilating into lactiferous sinuses beneath the areola. Automatic detection of abnormalities in these structures is facilitated by advanced imaging technologies such as digital mammography and magnetic resonance imaging (MRI), which can visualize and analyze these anatomical details in depth.

Adipose and connective tissue, which play a key role in breast shape and volume, vary from person to person, influenced by genetic and hormonal factors. Cooper's ligaments, bands of fibrous connective tissue, support the shape of the breast but can stretch over time, contributing to sagging. This connective tissue is also a potential site of tumor development, making its visualization essential for accurate diagnosis [14].

The nipple and areola, key areas for cancer detection, contain several galactophore duct openings for milk excretion. The areola contains Montgomery glands, secreting a lubricating and protective substance. Automatic diagnostic systems must be able to detect abnormalities in these areas, such as nipple retractions or abnormal discharge, which may be signs of cancer [15].

The breast anatomy is a multifunctional and dynamic structure essential for the automatic diagnosis of breast cancer. Advanced imaging and analysis technologies, coupled with a detailed understanding of breast anatomy, enable accurate and early detection and diagnosis of breast pathologies, facilitating effective management and appropriate surgical interventions.



**Figure 1.1 :** Breast anatomy.

## 2.2 Breast abnormalities

The sections below present a detailed overview of breast cancer abnormalities including masses, calcifications, architectural distortions, and asymmetries.

### 2.2.1 Microcalcifications

Breast microcalcifications are tiny calcium deposits that form in breast tissue. These calcifications are usually too small to be felt on physical examination. However, they can be detected on mammograms, where they appear as small bright white spots or dots less than a millimeter in size. Although the presence of microcalcifications is generally benign, certain morphologies or groupings may indicate underlying abnormalities, such as early stages of breast cancer. Consequently, identifying and analyzing microcalcifications plays a critical role in breast cancer screening and diagnosis [16]. Radiologists carefully assess the size, shape, and distribution of mammograms to determine whether further investigation, such as biopsy, is required to rule out malignancy.

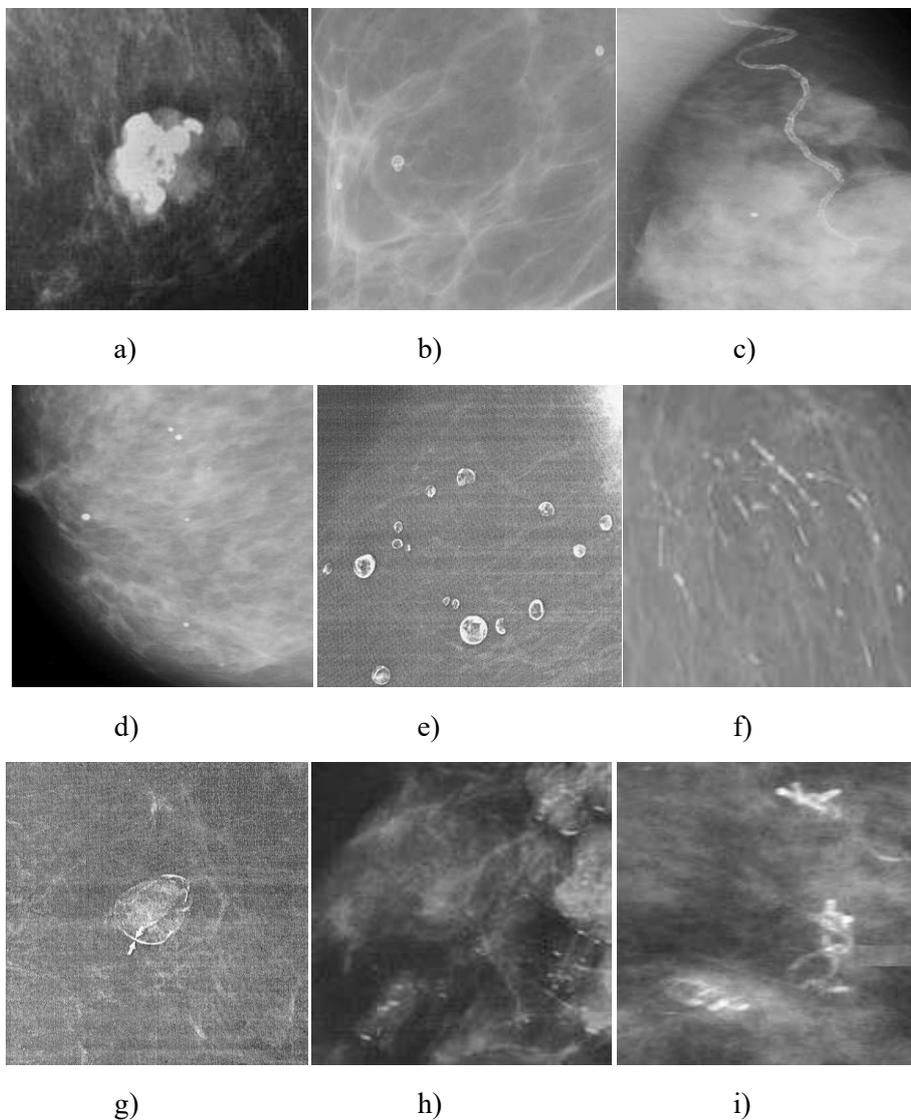
#### 2.2.1.1 Morphology

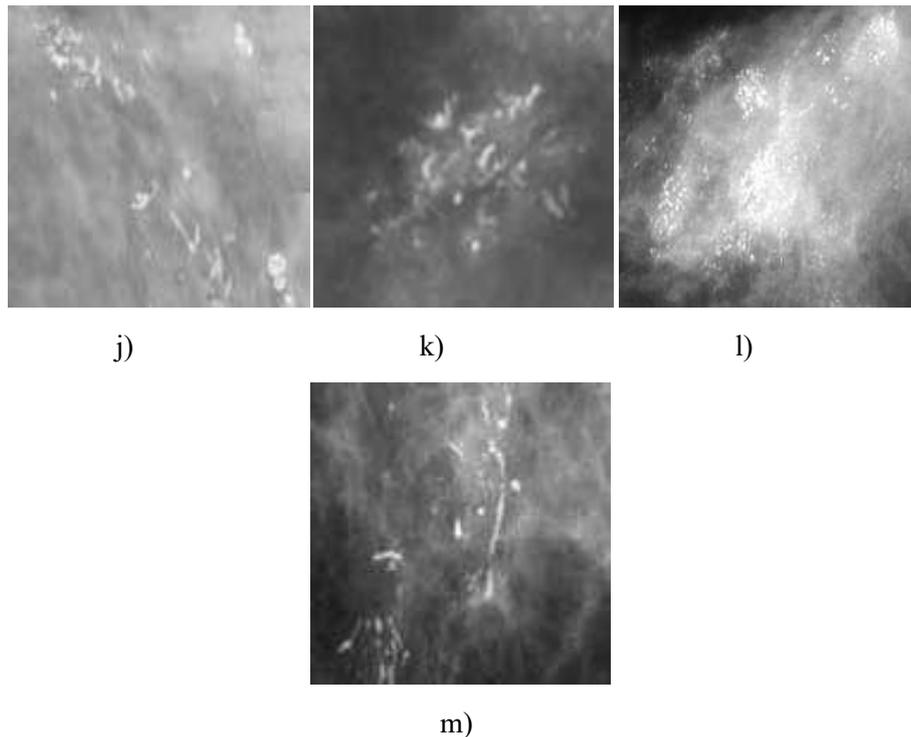
The morphology of microcalcifications is essential in the automatic diagnosis of breast anomalies, as it enables benign processes to be differentiated from malignant lesions with greater precision. Artificial intelligence (AI) and computed radiology systems rely on morphological analysis to identify suspicious shapes, such as fine and polymorphic or linear and branched microcalcifications, often associated with high cancer risk. Incorporating morphological criteria into CADx improves the performance of these systems and reduces the false positive rate, providing radiologists with a valuable aid for more accurate and rapid

diagnosis. In what follows, the different types of microcalcifications are detailed with illustrations in Figure 1.2 [17].

**a)** Coarse or coralliform microcalcifications: large calcifications (2 to 3 mm) with an irregular shape that may suggest more significant anomalies.

**b)** Cutaneous or dermal microcalcifications: They generally present no diagnostic difficulties on mammography, appearing as small, round, annular structures with clear centers. They are mainly located in the sub mammary, axillary, areolar, and parasternal regions, where subcutaneous glands are more abundant. If their cutaneous origin remains uncertain, additional incidences tangential to the skin may be recommended to clarify their location.





**Figure 1.2 :** The different types of microcalcifications: a) coarse or coralliform, b) cutaneous or dermal, c) vascular, d) round, e) clear center, f) rod like, g) eggshell or parietal, h) milk-calcium-like, i) suture-like, j) dystrophic, k) amorphous or indistinct, l) polymorphic fine, m) branched linear.

**c) Vascular microcalcifications:** are more frequent with age and predominate along the external mammary vessels. They may be a marker of atherosclerosis and coronary risk. They are easily recognized when they appear as long double-line patterns along the vessels, particularly in diabetic patients. Due to their linear distribution, they are more difficult to interpret when they are discontinuous at the beginning of the calcification process. In this case, histological verification may be necessary.

**d) Round microcalcifications:** These are round in shape and vary in size. Smaller than 0.5 mm, they are called punctiform or powdery. Depending on their size and distribution, these round microcalcifications may be associated with benign changes or require closer monitoring.

**e) Microcalcifications with clear centers:** Microcalcifications with clear centers are characterized by a distinctive shape with a lighter central core surrounded by a denser zone, often appearing as small halos or round structures. Their size generally varies but is often small, typically of the order of a few millimeters. This configuration may suggest benign processes, such as inflammatory changes or fibro-adenomas.

**f) Rod like microcalcifications** are elongated, narrow calcifications that resemble rods. Their particular shape may suggest pathological processes, such as precancerous changes or calcifications associated with benign lesions.

**g) Eggshell or parietal microcalcifications:** Eggshell or parietal microcalcifications are distinguished by their distinctive shape, where calcifications cluster in concentric layers around a center, mimicking the appearance of an eggshell. In terms of size, they are generally small, often of the order of a few millimeters, and their layered arrangement can vary in thickness and density. This morphology can help identify benign lesions such as calcified cysts.

**h) Calcium-milk microcalcifications:** Calcium-milk microcalcifications are characteristically granular or small, round particles, often with a dense, homogeneous distribution, resembling calcium milk in their appearance, similar to droplets or crystals. In terms of size, they are generally small, typically less than 1 mm. Their morphology is often associated with benign changes, such as dystrophic calcifications in fibro-adipal lesions or cystic changes.

**i) Suture microcalcifications:** These appear as small linear or filamentous calcifications along suture paths or in areas of scarring. These microcalcifications may appear on post-operative mammograms. They are generally small, often of the order of a few millimeters. Although benign, they require attention to avoid confusion with pathological abnormalities.

**j) Dystrophic microcalcifications** are large, irregular calcifications that usually form in damaged or scarred breast tissue. They often appear after surgery or radiotherapy or in areas of fat necrosis. In terms of size, they are generally larger than other types of microcalcification and vary in shape.

**k) Amorphous or indistinct microcalcifications:** Amorphous or indistinct microcalcifications are small, often less than 0.5 mm in size, with blurred outlines and an irregular shape. They lack a well-defined structure, making them difficult to characterize accurately. Because of their uncertain appearance, these microcalcifications may be associated with an increased risk of malignancy.

**l) Fine, polymorphic microcalcifications:** are small, irregular calcifications that vary in shape and size. They appear as fine lines or tiny dots, often with varied and complex contours. This variability in appearance makes them particularly suspicious, as they are frequently associated with malignant or precancerous lesions, such as ductal carcinoma in situ (DCIS).

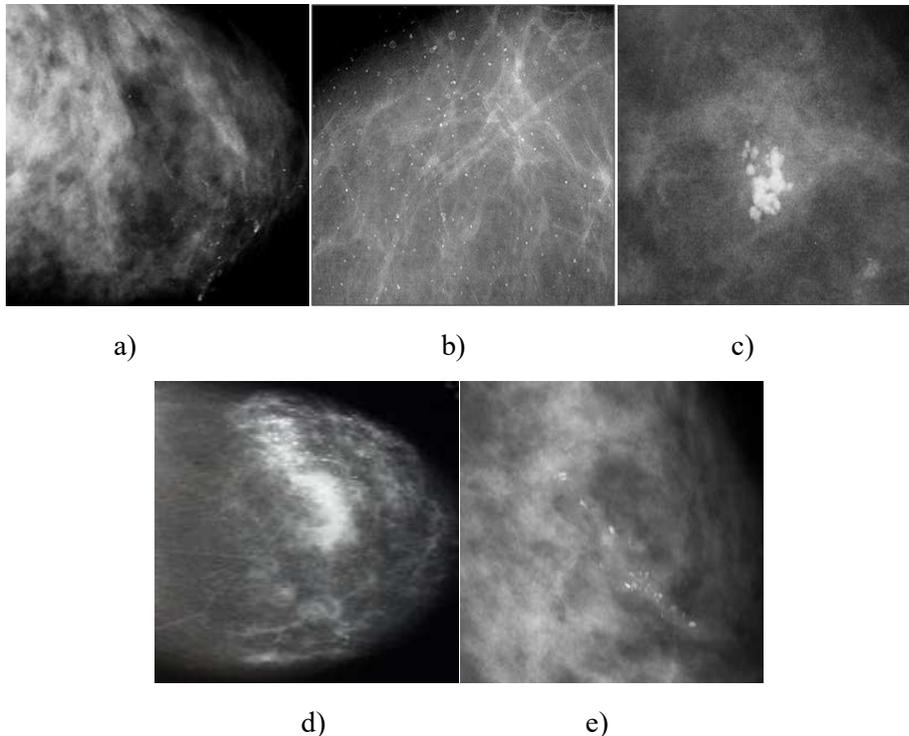
**m) Linear and branched microcalcifications:** Linear and branched microcalcifications are small calcifications that appear as straight or curved lines, often with branch-like extensions, evoking a treelike structure. Their arrangement along the mammary ducts is of particular concern, as they are frequently associated with malignant lesions.

#### **2.2.1.2 Distribution**

The distribution of microcalcifications is essential for their classification and diagnosis. Indeed, some distributions, such as clustered or linear microcalcifications, are more often associated with malignant lesions, while others, such as diffuse or scattered distributions, are

generally benign. Figure 1.3 shows the different types of distribution. Here are the main categories of distribution [18]:

- a) Diffuse distribution: Microcalcifications are randomly scattered throughout the breast, with no specific grouping. This distribution is generally benign and often associated with nonmalignant conditions such as skin calcifications or calcifications associated with necrotic tissue.
- b) Regional distribution: Microcalcifications are clustered over a wide breast area but are not confined to a single lobule or ductal segment. This distribution may be benign or suspicious, depending on the clinical context and the morphology of the calcifications.
- c) Clustered or clustered distribution: Microcalcifications are concentrated in a specific area, usually involving a single lobule or a small group of lobules. This distribution is of greater concern, as it may be associated with malignant lesions.
- d) Linear distribution: Microcalcifications follow a linear or ductal pattern, suggesting they are in the milk ducts. This distribution is often suspicious, as it may indicate an intraductal pathological process often linked to breast cancers.
- e) Segmental distribution: Microcalcifications are present in an anatomical segment of the breast, corresponding to a lobe or sector of the breast. This distribution is also worrying, as it may reflect a pathological process extending along a ductal segment, often associated with malignant tumors.



**Figure 1.3 :** The various distributions of microcalcifications: a) Segmental, b) Diffuse or scattered, c) Clustered, d) Regional, and e) Linear.

Each type of distribution gives clues to the nature of the microcalcifications and helps radiologists assess the risk of malignancy, guiding clinical decisions regarding biopsies and other additional investigations

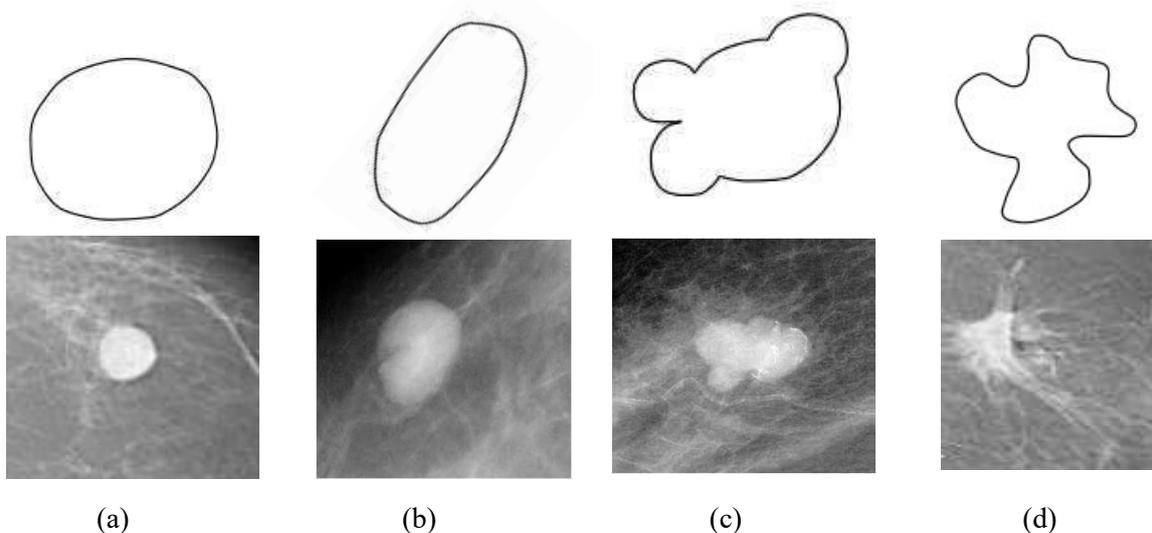
## 2.2.2 Masses

A breast mass is a palpable or visible abnormality in the breast tissue, usually seen as a nodule or lump. It can be detected by self-examination, clinical examination, or imaging techniques such as mammography, ultrasound, or MRI. Mammary masses can have various origins and natures, ranging from benign cysts or fibro-adenomas to malignant tumors, such as breast cancer. Their assessment is based on size, shape, contours, and density [19].

### 2.2.2.1 The Form

BI-RADS classifies breast masses according to their shape into four main categories:

- a) Round mass: Masses with a round or almost round shape are generally considered less suspicious. They are often associated with benign lesions such as cysts or fibroadenomas (figure 1.4 a).
- b) Oval mass: An oval or elliptical mass is also often benign. It may indicate a cyst or fibroadenoma, but whether it is benign or malignant will depend on other features such as contours and density (Figure 1.4 b).
- c) Lobular mass: Lobular breast cancer is a particular form of cancer that originates in the cells of the breast lobules with small undulations (figure 1.4 c).
- d) Irregular mass: Irregularly shaped masses are of greater concern, as they are more often associated with malignant lesions. An irregular shape may indicate a tumor whose contours are not well defined, suggesting a potential invasion of adjacent tissues (figure 1.4 d).

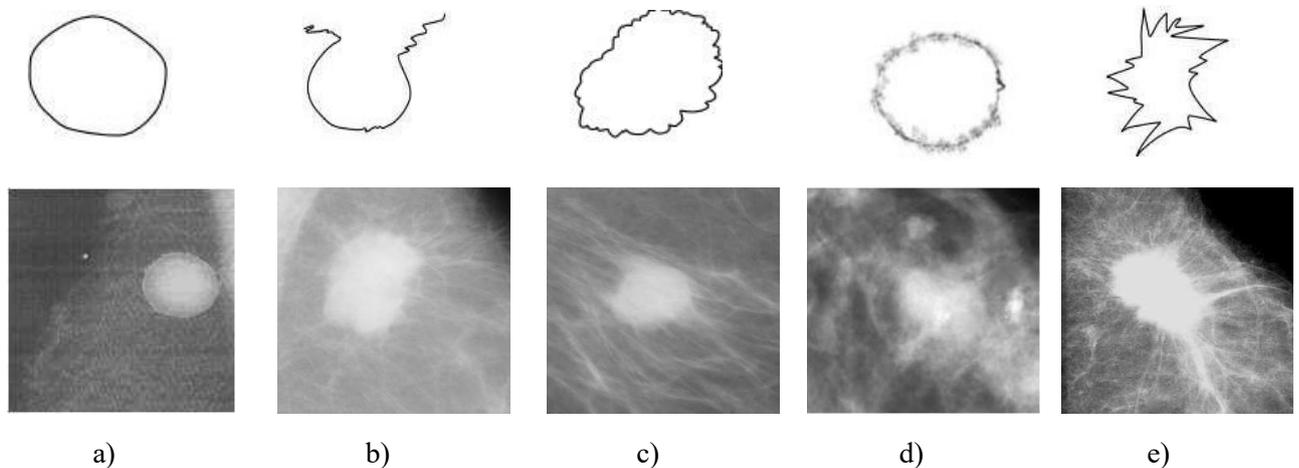


**Figure 1.4 :** The four shapes of a breast mass: a) Round, b) Oval, c) Lobulated, and d) Irregular.

### 2.2.2.2 The Contour

The contours of breast masses are important in assessing their nature. The characteristics of these contours can provide crucial clues as to whether a mass is likely to be malignant or benign [19].

- a) **Circumscribed contours:** Masses have delimited, well-defined margins, with a clear transition between the mass and the surrounding tissue. Circumscribed contours are often associated with benign lesions like simple cysts or fibroadenomas.
- b) **Micro-lobulated contours:** Masses have slightly lobulated margins with small projections or indentations. Micro-lobulated contours can be seen in benign and malignant lesions, often requiring further evaluation.
- c) **Masked contours:** The edges of the mass are difficult to define due to interaction with surrounding tissue, creating a blurred or indistinct appearance. Masked contours may be associated with malignant conditions
- d) **Spiculated contours:** Masses have edges that extend in "rays" or "spicules" into the surrounding tissue. Spiculated contours are often associated with malignant lesions.
- e) **Indistinct contours:** The edges of the mass are not clearly defined with surrounding tissues, which may indicate infiltration or interaction with neighboring tissues. This type of contour may be related to malignant conditions.



**Figure 1.5:** Different contour shapes for a mass: a) Circumscribed, b) Microlobulated, c) Masked, d) Indistinct, and e) Spiculated

### 2.2.2.3 Density

Breast density is not only an important technical factor in mammography interpretation, it is also a risk factor for breast cancer. The first to establish this link was Wolfe in 1976. Subsequent studies have questioned the link between density and breast cancer. Breast density stages are classified according to the American College of Radiology (ACR) BI-RADS (Breast Imaging Reporting and Data System). This classification describes breast tissue composition in four distinct categories, often called "breast density stages". These stages reflect the relative

proportion of glandular and fibrous tissue to adipose tissue in the breasts [20]. The four BI-RADS breast density stages are illustrated in Figure 1.6:

**Stage 1:** Breasts almost entirely fatty (less than 25% of the mammary gland).

- Description: The breasts mainly comprise adipose tissue, with very little glandular or fibrous tissue.
- Imaging: Mammograms are generally easier to interpret in this category, as adipose tissue appears in black, providing good contrast to detect abnormalities that appear in white.
- Risk: This density is associated with a relatively low risk of developing breast cancer and better visibility of abnormalities on mammograms.

**Stage 2:** Breasts with scattered areas of fibrous and glandular density (approximately 25-50% of the mammary gland).

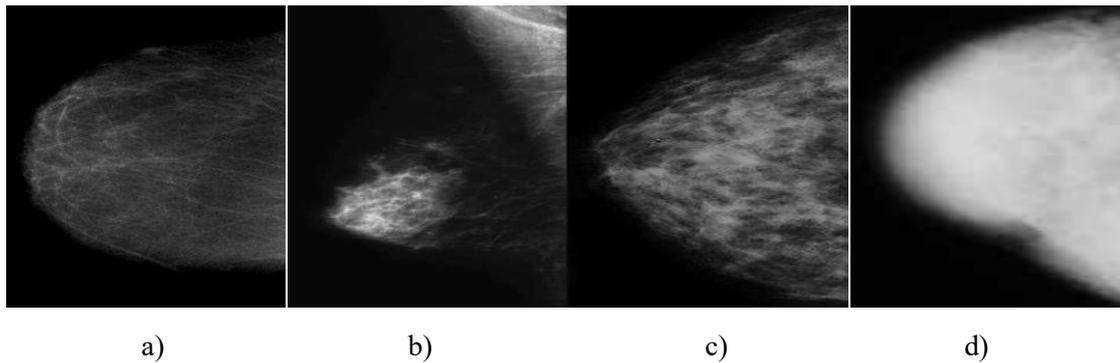
- Description: There are a few areas of glandular and fibrous tissue, but most of the tissue is adipose.
- Imaging: Abnormalities are generally clearly visible, although slightly less so than in stage 1.
- Risk: This density is also associated with a low risk of cancer, with good detection capacity during mammography.

**Stage 3:** Heterogeneously dense breasts (approximately 51-75% of the mammary gland).

- Description: A significant proportion of the breast comprises glandular and fibrous tissue, making the breasts dense overall.
- Imaging: This increased density can mask certain anomalies, making mammograms more difficult to interpret.
- Risk: Women with heterogeneously dense breasts have a moderate risk of breast cancer, and density can reduce the sensitivity of mammograms.

**Stage 4:** Extremely dense breasts (more than 75% of the mammary gland).

- Description: Most breast tissue is dense, meaning that the breasts are mainly composed of glandular and fibrous tissue, with little adipose tissue.
- Imaging: High tissue density can mask tumors, making detecting abnormalities more difficult.
- Risk: This category is associated with an increased risk of breast cancer. High density may also require additional imaging methods, such as MRI or ultrasound, for a more accurate assessment.



**Figure 1.6** : BIRADS classification of breast density: a) Density < 25%, b) Density between 25 and 50%, c) Density between 50 and 75% and d) Density > 75%.

### 2.2.3 Architectural distortions

Architectural distortion is a term used in breast imaging, particularly mammography, to describe an alteration in the normal architecture of the breast. It manifests as a change in the usual structure of breast tissue without a clearly defined visible or palpable mass [21]. Here is a more detailed description:

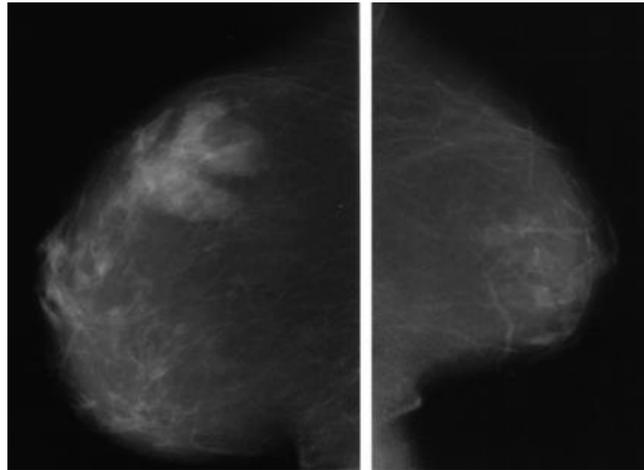
- a) Altered breast tissue lines: Breast tissue appears regular normally, with radial lines converging towards the nipple. In the case of architectural distortion, these lines are pulled, folded, or displaced abnormally, creating a disorganized appearance.
- b) Lack of a defined mass: Unlike a mass or tumor, architectural distortion is not a delineated entity. This means that there is no clear edge or distinct identifiable shape.
- c) Rays and spiculated lines: Rays or spiculated lines may be observed converging on a focal point but without a palpable or visible mass. This can appear as "dashes" or "spicules" emanating from a central region.
- d) Tissue shrinkage: Sometimes, architectural distortion can cause surrounding tissue to shrink or sag, giving the impression that the tissue is being "pulled" towards a particular point.



**Figure 1.7** : Example of an architectural distortion.

### 2.2.4 Breast asymmetry

Breast asymmetry, or asymmetry in breast size, shape, or density, maybe a normal variation, but it can also be considered an anomaly in certain contexts (Figure 1.8) [22].



**Figure 1.8 :** Left and right profile mammogram showing asymmetry of shape.

**a) Normal breast asymmetry**

- **Physiological variation:** Many women have a certain natural asymmetry between their breasts, where one breast is slightly larger or differently shaped. This asymmetry is generally benign and does not cause health problems.
- **Natural evolution:** Breast size and shape can change over time due to hormonal fluctuations, pregnancy, breastfeeding, or aging factors. These changes can sometimes accentuate an asymmetry already present.

**b) Breast asymmetry as an anomaly**

- **Developed asymmetry:** If breast asymmetry appears suddenly or develops rapidly, it may be considered an abnormality requiring further evaluation. A recent change in the size or shape of one breast relative to the other may indicate an underlying pathology, such as a tumor or other breast condition.
- **Focal asymmetry on mammography:** Focal asymmetry on mammography, where one area of the breast is denser than the other, can be a sign of pathology. If asymmetry is seen for the first time on mammography, particularly if it is associated with other abnormalities such as masses or microcalcifications, it could indicate the presence of breast cancer or another breast condition.
- **Asymmetry of density:** A marked difference in breast tissue density between the two breasts may also require investigation, as it can sometimes mask abnormalities or indicate an underlying lesion.

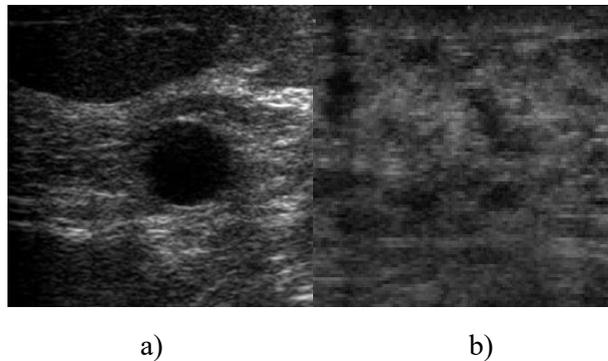
### **3. Screening methods**

Medical diagnosis relies on various imaging tools, each exploiting different physical properties to visualize organs. In the field of breast cancer, ultrasound (ultrasound imaging), MRI (Magnetic Resonance Imaging), and mammography (X-ray imaging) are commonly used to detect and characterize suspicious lesions. Medical breast imaging combines different techniques, such as ultrasound, MRI, and mammography, to provide a complete and accurate

view of breast tissue. This multimodal approach is essential for reliable diagnosis and optimal patient follow-up [23]. In what follows, we present the different techniques used and their characteristics.

### 3.1 Ultrasound

Ultrasound, or ultrasound imaging, is a medical imaging technique that uses high-frequency sound waves to produce real-time images of the body's internal organs. These images are then displayed on a screen, enabling doctors to examine the anatomical structures in detail. The principle is that a probe emits (Ultrasound emission) high-frequency sound waves that penetrate the body. The various tissue interfaces reflect These waves (Ultrasound reflection), creating echoes. In the final imaging stage, the probe captures the echoes and then converts them into a real-time image on a screen. Figure 1.9 shows two ultrasound images, a) with breast mass and b) normal breast with no abnormalities.



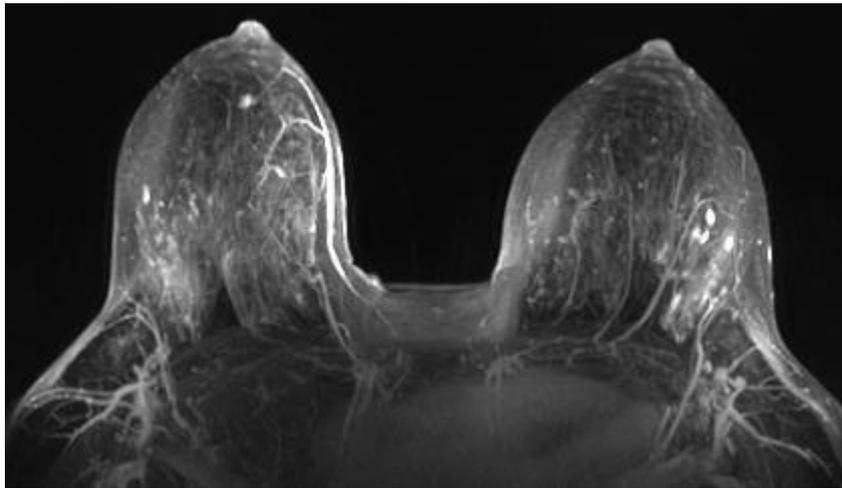
**Figure 1.9 :** Examples of breast ultrasound: a) ultrasound with abnormality (mass), b) normal breast.

Ultrasound offers several advantages, including real-time visualization, allowing for the observation of organ and tissue movements, which is particularly useful for assessing the nature of a lesion. It is a non-invasive and painless procedure that does not expose patients to X-rays, making it a safer option, especially for young women and pregnant women. Additionally, ultrasound is more affordable compared to other imaging tests and serves as a valuable complement to mammography, particularly in distinguishing cysts from solid tumors. However, Ultrasound has some disadvantages, including limited penetration, making it less effective for detecting deep breast lesions. It also has limited sensitivity to microcalcifications, which are often an early sign of breast cancer. Additionally, the quality of the examination depends on the operator's experience and skill, and visualization may be challenging in women with dense breasts.

In short, ultrasound is an excellent complement to mammography but should not be replaced. It is particularly useful for assessing the nature of a lesion that has already been identified but is less effective for detecting microcalcifications. For optimal breast cancer screening, mammography remains the gold standard.

### 3.2 Magnetic resonance imaging

Magnetic resonance imaging (MRI) is a medical imaging technique that exploits the magnetic properties of the hydrogen atoms in our bodies. These atoms align in a specific orientation in a powerful magnetic field. Radio waves disrupt this alignment; the atoms release this energy through radio signals that the device picks up. By analyzing these signals, a computer reconstructs detailed images of organs and tissues, enabling the detection of any abnormalities. The different behavior of hydrogen atoms in different tissues enables contrasting images to be created and anatomical structures to be visualized with great precision without requiring X-rays.



**Figure 1.10** : Example of breast MRI bi-sein.

MRI offers several advantages, including high precision with excellent spatial resolution, allowing for detailed visualization of fine anatomical structures. It is a non-invasive, non-ionizing technique that does not use X-rays, making it safer for patients. Additionally, MRI provides multiple contrast images, aiding in the differentiation between healthy and diseased tissue. Moreover, it usually requires no special preparation, making it more convenient compared to other imaging tests.

MRI has several disadvantages, including a longer examination time compared to ultrasound or X-ray, which may cause discomfort for some patients. It is also more expensive than other imaging methods and has contraindications for individuals with metal implants or objects in their bodies. Additionally, the confined space can be problematic for patients with claustrophobia, and the loud noise during the procedure may be unpleasant. In some cases, a contrast medium injection is required, posing a minimal risk of allergic reaction. In summary, MRI is a powerful medical imaging technique that offers many advantages regarding accuracy and safety. However, it also presents certain limitations, notably cost and patient comfort. The choice of MRI will depend on the nature of the examination to be carried out and any contraindications.

### 3.3 Tomosynthesis

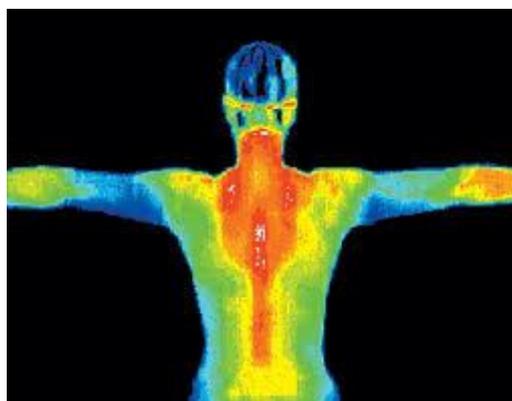
Tomosynthesis, or 3D mammography, is a new imaging technology used for breast examination. Unlike conventional mammography, where the X-ray emitting tube remains fixed, the tomosynthesis device is equipped with a mobile X-ray tube. This tube moves in an arc around the breast, taking images from different angles. A computer then processes all these images, combining them to create a detailed three-dimensional (3D) representation of the mammary gland. This approach enables breast tissue to be visualized in thin layers, improving the detection of abnormalities, particularly in dense breasts where structures can overlap and mask lesions in traditional 2D mammography.

Tomosynthesis has better differentiation between normal tissue and suspicious areas, thus reducing the false positive rate and the need for further tests. Enabling radiologists to scan the breast layer by layer virtually improves the detection of small masses and architectural distortions that might go unnoticed in conventional mammography.

This technology is increasingly used in breast cancer screening and diagnosis, offering a more accurate option for assessing patients, particularly those with dense breasts or at high risk of cancer.

### 3.4 Thermography

Thermography is a non-invasive imaging modality used in breast cancer detection that captures the infrared radiation (heat) emitted from the skin surface. Cancerous tissues typically exhibit higher metabolic activity and increased blood flow, resulting in localized temperature elevations that can be visualized as thermal anomalies. By using infrared cameras, thermographic imaging can highlight abnormal heat patterns, asymmetries, or vascular changes associated with potential malignancies. Unlike mammography, thermography does not use ionizing radiation and is painless, making it suitable for younger women or those with dense breast tissue. However, its clinical utility is limited due to lower sensitivity and specificity compared to conventional imaging techniques. As such, thermography is generally considered a supplementary tool rather than a primary diagnostic method, often used in conjunction with mammography or ultrasound for improved diagnostic accuracy. Figure 1.11 shows an example of thermography image.



**Figure 1.11:** Thermography image.

### 3.5 Positron Emission Tomography

Positron Emission Tomography (PET) is an advanced nuclear imaging modality used in breast cancer diagnosis, staging, and monitoring. It provides functional imaging by detecting metabolic activity within tissues. In PET imaging, a small amount of radioactive tracer commonly fluorodeoxyglucose (FDG), is injected into the body. Cancer cells typically have a higher rate of glucose metabolism than normal cells, so they absorb more of the tracer. The PET scanner then detects the gamma rays emitted as the tracer decays, producing images that reflect metabolic activity rather than just anatomical structure. In breast cancer, PET is particularly useful for detecting distant metastases, evaluating lymph node involvement, and monitoring treatment response. It is often combined with CT in a hybrid PET/CT scan, which merges functional and anatomical data for more accurate localization of abnormalities. Although highly effective in detecting active disease, PET is not typically used for routine breast cancer screening due to its high cost, limited resolution for small lesions, and exposure to ionizing radiation. Instead, it plays a key role in advanced-stage assessment and therapy planning.

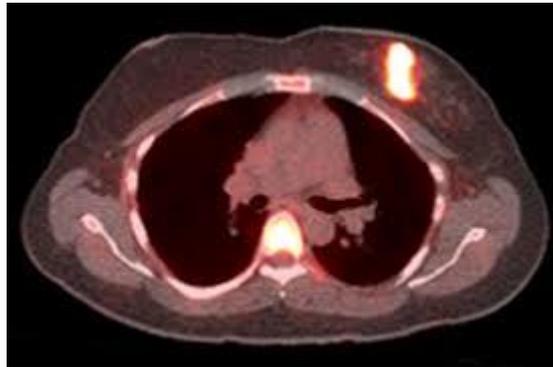
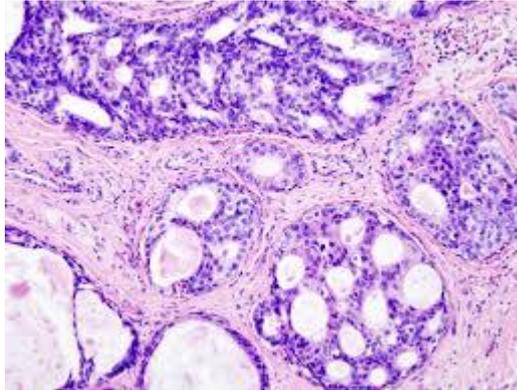


Figure 1.12 : Positron Emission Tomography image.

### 3.6 Histopathology

Histopathological analysis is the gold standard for definitive breast cancer diagnosis. It involves the microscopic examination of breast tissue samples, typically obtained via biopsy to determine the presence, type, and characteristics of cancerous cells. In this process, the tissue is first fixed, sectioned, and stained (commonly with Hematoxylin and Eosin) to highlight cellular structures. Then, a pathologist examines the samples under a microscope to assess features such as cell morphology, nuclear atypia, mitotic activity, and tissue architecture. Histopathology not only confirms malignancy but also provides critical details like tumor grade, invasiveness, margins, and lymphovascular invasion. While imaging modalities suggest the presence of abnormalities, histopathological analysis provides the final, conclusive diagnosis and is essential for accurate classification and staging of breast cancer.



**Figure 1.13:** An example of histopathological image.

## 3.7 Mammography

Mammography is a medical imaging technique that uses low-dose X-rays to examine breast tissue. It is mainly used for breast cancer screening and diagnosis. Mammography is the reference screening tool for the early detection of breast abnormalities, including masses, microcalcifications, and other potential signs of cancer. The principle of mammography is based on using X-rays to produce two-dimensional images of the breasts. During the examination, the breast is compressed between two plates to spread the breast tissue, thereby reducing the required radiation dose and improving image quality. The X-ray tube emits radiation that passes through the breast and is captured by a detector (either photographic film or a digital detector), creating an image that shows the different densities of breast tissue. Denser structures, such as tumor masses or microcalcifications, appear in white, while less dense adipose tissue appears in darker gray. This difference, in contrast, enables radiologists to detect potential abnormalities.

### 3.7.1 Mammography incidences

In mammography, incidences refer to the different views or projections used to visualize the breasts. These views are essential for obtaining a complete and detailed representation of breast tissue, enabling any abnormalities to be detected. The two basic views, mediolateral oblique (MLO) and craniocaudal (CC) are generally supplemented by additional views to clarify areas of interest, if necessary.

**a) Medio-lateral oblique incidence (MLO):** The MLO view is the most commonly used in mammography. It is obtained by placing the breast on the compression plate at an angle of 30 to 60 degrees to the vertical axis, thus capturing an image of the upper part of the breast up to the armpit. This incidence covers most of the breast tissue, including Spence's tail, which extends into the armpit. Also, the oblique angle allows for the good visualization of deep breast structures. The only drawback of MLO is image quality, which can vary according to the radiologist's technique, particularly in axillary coverage.

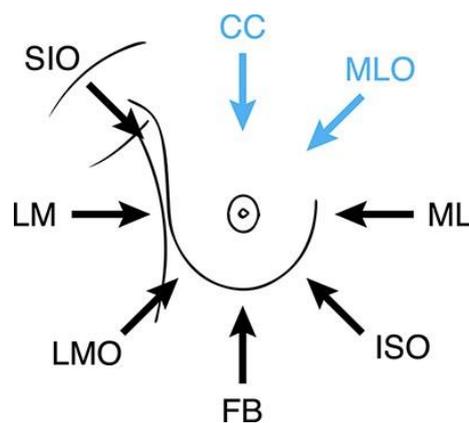
**b) Cranio-caudal incidence (CC):** The CC incidence is a top-down view, with the breast compressed horizontally between two plates. The resulting image shows a cross-section of the breast. The CC view is complementary to the MLO view, enabling anomalies to be localized in

relation to the medial or lateral position of the breast. This view enables a symmetrical comparison between the two breasts. The limited coverage of this view prevents radiologists from reaching the upper or lateral regions of the breast, particularly areas close to the armpit.

c) Additional impacts

- Strict profile incidence (ML: mediolateral or LM: lateromedial): A view taken strictly from the side, either from the outside inwards (mediolateral) or from the inside outwards (latero-medial). This view complements the MLO and CC incidences for precise localization of a suspicious lesion in the horizontal plane.
- Magnified Incidence: Uses specific compression support and a shorter focal length to magnify a particular breast area. Particularly useful for examining microcalcifications or small lesions in greater detail.
- Spot compression incidence: Its principle is based on localized compression on a specific breast area to improve clarity and reduce tissue overlay. It helps to define better anomalies detected during standard views.
- Tangential incidence: A view that tangents the breast to visualize superficial lesions. Used to evaluate cutaneous or subcutaneous abnormalities.
- Axillary incision (Cleavage or Axillary): View obtained to better visualize the armpit and Spence's tail. Used to evaluate lymph nodes or abnormalities close to the axilla.

Combining these incidences provides a complete representation of the breast, making detecting abnormalities in different parts of the breast tissue easier. By adjusting the angle and direction of the image, radiologists can locate lesions more precisely, assessing their size, shape, and nature (solid or cystic). This information is essential for accurate interpretation of mammograms and for guiding subsequent diagnostic decisions, such as the need for a biopsy or closer follow-up. Figure 1.14 illustrates the different types of incidences.



**Figure 1.14 :** The different mammography views, with MLO and CC being the most commonly used.

### 3.7.2 Screening and diagnostic mammography

Screening mammography is a systematic examination offered to symptom-free women to identify subtle abnormalities in breast tissue that may escape clinical palpation. Its primary role is the early detection of breast cancers at a stage when they are generally smaller and easier to treat. This examination is conducted according to a rigorous protocol, enabling systematic comparison of the two breasts and in-depth analysis by specialized radiologists.

Diagnostic mammography, on the other hand, is requested in a specific clinical context when particular signs or symptoms lead to the suspicion of the presence of a breast lesion. These may include a palpable mass, pain, a change in the skin or nipple, or an abnormality detected during a screening examination. Diagnostic mammography aims to characterize the lesion observed, to determine whether it is benign or malignant, and to guide further investigations if necessary.

### 3.8 Evaluation of different modalities

Table 1.1 summarizes the key characteristics of various breast cancer screening and diagnostic modalities [24]. Mammography remains the standard for population screening due to its efficiency and accessibility, especially in detecting masses and microcalcifications.

Modality	Principle	Resolution	Cost	Use Case	Strengths	Limitations
<b>Mammography</b>	X-ray imaging	High (for masses & calcifications)	Low–Medium	Primary screening, especially in women >40	Detects microcalcifications; widely available	Less effective in dense breasts
<b>Ultrasound</b>	Sound waves	Moderate (soft tissue)	Low	Adjunct to mammography; dense breasts	No radiation; good for cyst vs. solid differentiation	Operator-dependent; less effective for microcalcifications
<b>MRI</b>	Magnetic fields and radio waves	Very high	High	High-risk screening, staging	Excellent soft tissue contrast; sensitive	Expensive; may cause false positives
<b>Thermography</b>	Infrared heat detection	Low	Low	Experimental/supplementary	Non-invasive, radiation-free	Low specificity and sensitivity

<b>PET / PET-CT</b>	Radioactive tracer detects metabolism	Moderate–High (functional)	Very High	Staging, recurrence, therapy response	Functional imaging; detects metastasis	Not for screening; expensive; radiation exposure
<b>Histopathology</b>	Microscopic tissue analysis (biopsy)	Cellular-level (definitive)	Medium	Diagnostic confirmation	Gold standard; precise tumor characterization	Invasive; requires tissue sampling

**Table 1.1** : Comparison between different screening modalities.

#### 4. The BI-RADS Classification

The BI-RADS system was designed to standardize mammography reporting and facilitate clinical decision-making [25]. Developed by the American College of Radiology (ACR), it is used in many countries and is a benchmark in managing breast cancer screening and diagnosis. BI-RADS, first published in 2013, standardizes the evaluation of abnormalities detected through three types of imaging, namely mammography, ultrasound, and MRI. BI-RADS classification plays a central role in CADx systems, as these systems are often designed to align with BI-RADS criteria to assist radiologists in classifying detected abnormalities and reduce variations in image interpretation. The latest edition of BI-RADS is version 5, published in 2013, and comprises 5 classes; the following table illustrates the different categories of this standard.

<b>Classes</b>	<b>Interpretation</b>	<b>Risk</b>	<b>CAT</b>
BI - RADS 0	Incomplete investigation		Reviews
BI - RADS I	Normal image	0%	Screening
BI - RADS II	Benign lesion	0%	Screening
BI - RADS III	Probably benign lesion	< 2%	Follow-up
BI - RADS IV	Suspected lesion requiring biopsy	> 2 et < 90%	Biopsy
BI - RADS V	Lesion highly suggestive of cancer	> 90%	Biopsy

**Table 1. 2** : Interpretation and action to be taken (CAT) according to BI-RADS classes.

#### 5. Computer aided diagnosis and detection systems

Diagnosis is a critical step in the management of breast cancer, as it significantly influences the patient's prognosis. When detected at an early stage, breast cancer offers considerably

higher chances of successful treatment and survival compared to cases identified at a more advanced stage. As a result, extensive research has been dedicated to developing advanced screening and diagnostic tools aimed at improving the early detection and accurate classification of breast tumors. In this context, Computer-Aided Detection (CADe) and Computer-Aided Diagnosis (CADx) systems have emerged as valuable clinical tools, assisting radiologists in identifying and interpreting subtle patterns in breast images. With the advent of digital technologies, the integration of Artificial Intelligence (AI), particularly deep learning into these systems has opened new horizons. AI-driven models can automatically extract and analyze complex imaging features, offering promising improvements in diagnostic precision, sensitivity, and workflow efficiency in breast cancer detection and diagnosis.

## **6. Conclusion**

In conclusion, this chapter has provided a comprehensive overview of breast cancer, covering its various aspects, from breast anatomy to the most common abnormalities, risk factors, and screening methods. The anatomical structure of the breast, comprised of lobules, ducts, adipose tissue, and connective fibers, is crucial to understanding how and where abnormalities can develop. Particular attention was paid to analyzing the specificities of observed breast pathologies, such as masses, calcifications, and architectural distortions, which are important indicators of the potential presence of malignant or benign tumors, often detected through medical imaging. The chapter also highlighted the importance of the BI-RADS classification, an essential tool for assessing and reporting these anomalies according to their level of risk. Thanks to this system, radiologists can standardize diagnoses and better guide clinical decisions, particularly concerning lesions requiring surveillance or biopsy. This chapter aims to provide an introduction computer aided diagnosis and detection systems (CADx/CADe), that will be detailed in chapter 2.

## Chapter 2: CADx systems for breast cancer

### 1. Introduction

Several imaging modalities for the detection and diagnosis of breast cancer are used and plays a crucial role in reducing mortality through early identification of malignancies. However, the essential challenges associated with breast images interpretation persist. A primary difficulty lies in the visual similarities between early signs of breast cancer and typical anatomical structures within breast tissue, which often complicates the differentiation of benign from malignant findings. This issue is further aggravated by dense breast tissue, overlapping structures, and subtle morphological variations that may obscure pathological changes or mimic malignant lesions. Mammography is frequently supplemented with other imaging modalities, notably ultrasound and magnetic resonance imaging (MRI). Ultrasound is particularly valuable for distinguishing between cystic and solid masses, while MRI is renowned for its high sensitivity in detecting abnormalities, especially in dense breasts or high-risk populations. These approaches improve diagnostic accuracy but do not entirely resolve the complexities in breast imaging interpretation [34].

In response to these challenges, computer-aided detection (CADe) and computer-aided diagnostic (CADx) systems have been developed to support radiologists in detecting and characterizing breast abnormalities and considered as second reader. CADe systems are primarily designed to identify regions of interest, such as potential lesions or microcalcifications [27]. In contrast, CADx systems aim to provide a more nuanced assessment by analyzing the probability of malignancy based on features such as lesion shape, size, and texture. Over the past two decades, significant advancements have been made in developing CADe and CADx models, primarily driven by innovations in artificial intelligence (AI) and machine learning. Deep learning algorithms, in particular, have enabled the analysis of large-scale imaging datasets with unprecedented precision. By training on annotated datasets comprising mammograms, these models can detect intricate patterns and features that might escape human observers. Consequently, these technologies promise to enhance diagnostic accuracy, reduce the time required for image review, and mitigate the cognitive load on radiologists.

Despite their promise, the clinical integration of CAD systems is not without challenges. High false-positive rates remain a significant concern, leading to unnecessary biopsies, increased patient anxiety, and elevated healthcare costs. Moreover, the performance of these systems is contingent upon the quality and diversity of the datasets used for training, highlighting the importance of rigorous dataset curation and validation.

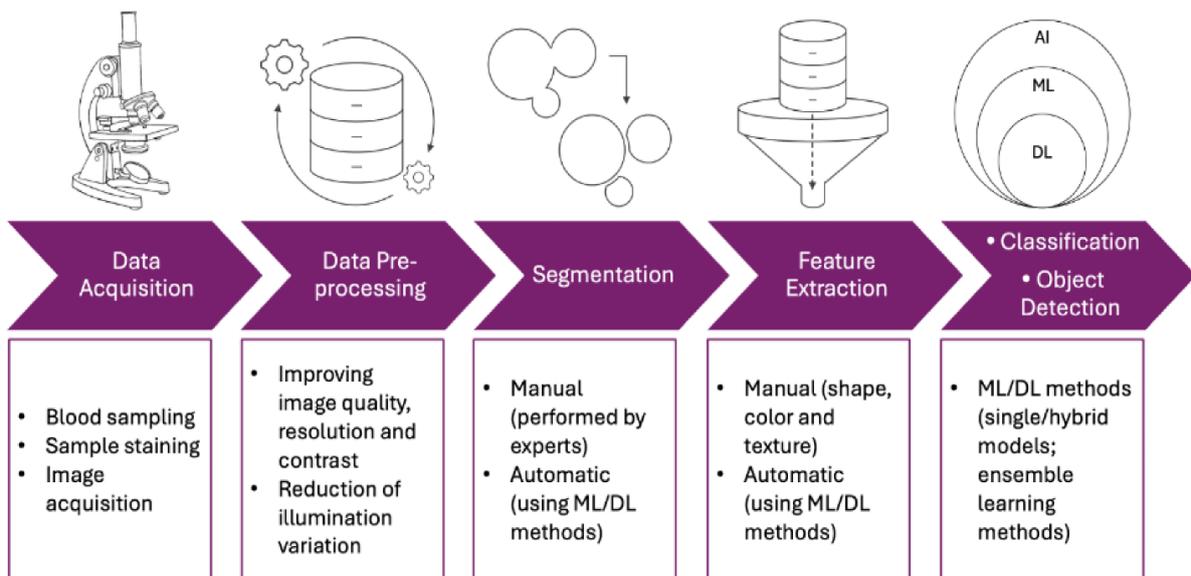
### 2. CADx systems

This section presents a detailed overview of computer-aided diagnosis systems, from the definition to the constituent stages of the system, including the methods and algorithms used in each phase.

## 2.1. Definition

Computer-aided diagnostic (CADx) systems are sophisticated computational platforms developed to support healthcare providers in evaluating and interpreting medical imaging data for diagnostic applications. These systems harness the power of artificial intelligence (AI), machine learning (ML), and advanced image processing algorithms to analyze medical images, such as mammograms, ultrasounds, or MRIs, and deliver detailed assessments of identified anomalies. While Computer-Aided Detection (CADe) systems are primarily tasked with locating and flagging regions of interest (ROIs) that may suggest potential abnormalities, CADx systems extend this functionality by offering a deeper analysis of these regions. They aim to characterize the nature of the abnormalities, such as their morphological and textural properties, and estimate their likelihood of being benign or malignant [27].

The five key steps in CADx systems are data acquisition, preprocessing, segmentation, feature extraction and selection, and classification as illustrated in Figure 2.1. First, data acquisition collects data from instruments like MRI or mammograms. After that, preprocessing enhances raw medical images by reducing noise (e.g., Gaussian filtering), correcting intensity inhomogeneities, and improving contrast to ensure consistent input for analysis. Then, segmentation isolates regions of interest (e.g., tumors, organs) using techniques like thresholding, region-growing, or deep learning models to delineate anatomical or pathological structures. The third step is feature extraction and selection, permitting quantifying and refining discriminative attributes (e.g., texture, shape, radiomic features) from segmented regions, employing methods like PCA or mutual information to retain diagnostically relevant data. Finally, classification applies machine/deep learning models (e.g., CNNs, SVM) to categorize findings into diagnostic classes (e.g., benign vs. malignant), providing probabilistic outputs to aid clinical decision-making. These steps collectively transform raw imaging data into actionable diagnostic insights. In the next section each step will be detailed [28].



**Figure 2.1 :**Workflow of Computer-Aided Diagnosis and Detection systems [178].

## 2.1 Data Acquisition

Data acquisition is a foundational step in developing Computer-Aided Diagnosis (CADx) systems for breast cancer, as it directly influences the system's performance and generalizability. High-quality, diverse, and representative datasets are essential for training and validating diagnostic algorithms effectively [24].

In medical CADx systems, data acquisition typically involves:

1. **Image Collection:** Medical images such as mammograms, ultrasounds, and MRIs are gathered from clinical sources or publicly available databases. For instance, the Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) provides annotated mammographic images for research purposes.
2. **Labeling and Annotation:** Expert radiologists annotate regions of interest (ROIs), classify findings (e.g., benign vs. malignant), and provide metadata like patient demographics. These annotations are crucial for supervised learning models.
3. **Preprocessing:** Raw images undergo preprocessing steps such as noise reduction, normalization, and contrast enhancement to standardize inputs and improve model performance.
4. **Data Balancing and Augmentation:** Medical datasets often exhibit class imbalance (e.g., fewer malignant cases). Techniques like data augmentation (rotations, flips) or synthetic data generation are applied to address this issue.
5. **Dataset Splitting:** The dataset is divided into training, validation, and testing sets to evaluate the system's accuracy and generalizability.

Effective data acquisition ensures that CADx systems are trained on high-quality, diverse, and clinically relevant data, enabling them to assist radiologists more accurately in diagnosing breast cancer.

## 2.2 Preprocessing

Preprocessing as first step, plays an indispensable role in Computer-Aided Diagnosis (CADx) systems for breast cancer, serving as the foundational step that significantly influences the accuracy and reliability of the diagnostic process. By enhancing the quality of breast images, preprocessing enables the CADx system to more effectively identify subtle and complex abnormalities, such as masses, microcalcifications, and architectural distortions, which are often early indicators of breast cancer. These abnormalities can be challenging to detect, due to raw images inherently low contrast, noise, and artifacts [29]. Preprocessing tasks address these challenges by reducing noise, enhancing contrast, removing artifacts, standardizing image properties, and transforming the raw data into a more interpretable and analyzable form. In simpler terms, preprocessing acts as a refining filter, ensuring the CADx system operates on more transparent, accurate, and consistent images. This refinement improves the system's ability to detect potential issues and reduces the likelihood of false positives and negatives, ultimately leading to more precise and reliable diagnoses. However, one of the most challenging aspects of preprocessing is ensuring that essential features of the images, which are crucial for CADx analysis, are not unintentionally removed or altered. Over-smoothing or excessive noise

reduction can obscure fine details, such as subtle microcalcifications or faint architectural distortions, critical for early breast cancer detection. Striking a balance between enhancing image clarity and preserving diagnostically relevant features is crucial to maintaining the effectiveness of CADx systems. By optimizing the input data while safeguarding essential information, preprocessing empowers CADx systems to support radiologists in making informed decisions, thereby playing a pivotal role in the early detection and diagnosis of breast cancer, which is critical for improving patient outcomes. Most important tasks of preprocessing can be summarized as follows:

### 2.2.1 Noise reduction

In medical imaging, noise refers to random variations in pixel intensity that can blur fine details and reduce the clarity of the image. This noise can come from several sources, such as electronic interference from the imaging equipment, how X-rays are absorbed by tissue or even the structure of overlapping breast tissues. Other factors, like patient movement or scatter radiation, can also add to the noise, making the image more challenging to interpret.

When there is too much noise, it can hide important features like tiny calcium deposits (microcalcifications), small lumps (masses), or changes in tissue structure (architectural distortions), all of which are crucial for detecting breast cancer. Preprocessing techniques are used to reduce the noise, like applying special filters or adjusting contrast. These methods improve image quality, allowing the CADx system to pick up on the key details for diagnosis, making it easier for doctors and radiologists to spot potential issues more accurately.

#### 2.2.1.1 Gaussian Filtering:

This technique smooths the image by averaging the pixel values around each point using a Gaussian kernel (a mathematical function that resembles a bell curve). The kernel helps to blur out the finer details of high-frequency noise, effectively reducing random variations while preserving larger structures like edges. It's particularly useful when the noise is spread across the image in a consistent, low-frequency pattern. The benefit of Gaussian filtering is that it keeps the image smooth without distorting important boundaries, which is crucial for maintaining the integrity of the tissue structures [30]. The kernel is based on the Gaussian function, which is defined as:

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (2.1)$$

Where:

- $G(x,y)$  is the value of the Gaussian function at the point  $(x,y)$ ,
- $\sigma$  is the standard deviation, which controls the width of the Gaussian bell curve.

The image  $I(x,y)$  is filtered by convolving it with the Gaussian kernel  $G(x,y)$ :

$$I_{filtered}(x, y) = \sum_{i=-k}^k \sum_{j=-k}^k I(x+i, y+j) \cdot G(i, j) \quad (2.2)$$

Where:

- $I(x,y)$  is the original image,
- $I_{\text{filtered}}(x,y)$  is the resulting filtered image,
- $k$  is the size of the kernel.

This equation represents the convolution of the image with the Gaussian kernel, which helps reduce noise while maintaining edges.

### 2.2.1.2. Median Filtering:

Median filtering works by replacing each pixel's value with the median value of its surrounding neighborhood [31]. It's particularly effective in removing "salt-and-pepper" noise, which appears as random black and white speckles scattered throughout the image. Unlike Gaussian filtering, which averages pixel values, median filtering focuses on removing extreme values (outliers) while preserving edges and other important structures. This technique is useful in cases where noise is highly irregular and doesn't follow a consistent pattern. If  $N(x,y)$  represents the set of pixel values in the neighborhood of pixel  $(x,y)$ , the median filter replaces the value of  $I(x,y)$  with the median value from this set:

$$I_{\text{filtered}}(x,y) = \text{median} ( N (x,y) ) \quad (2.3)$$

Where:

- $N(x,y)$  is the neighborhood of pixel  $(x,y)$ ,
- $I_{\text{filtered}}(x,y)$  is the new pixel value after applying the median filter.

In this context, the median of the neighborhood is calculated by sorting the pixel values and selecting the middle value, which effectively removes outliers like salt-and-pepper noise while preserving edges and other important structures.

### 2.2.1.3 Wavelet Transform:

The wavelet transform is a more advanced technique that breaks the image down into different frequency components. It separates the image into low-frequency (smooth areas) and high-frequency (detailed areas) components. Noise typically exists in the high-frequency part, so by selectively reducing or removing these high-frequency components while preserving the low-frequency ones, the image becomes cleaner without losing significant details [32]. This allows for highly targeted noise reduction, making wavelet transform particularly effective in preserving both fine structures (such as microcalcifications) and larger, important features (like masses or lesions) during the denoising process. The basic wavelet transform equation is:

$$W(x,y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} f(x',y') \cdot \varphi(x-x',y-y') dx'dy' \quad (2.4)$$

Where:

- $W(x,y)$  is the wavelet transform of the image  $f(x,y)$ ,
- $\varphi(x-x',y-y')$  is the wavelet function, which is typically a scaled and shifted version of a base wavelet.

After performing the transform, the image is split into approximation (low-frequency) and detail (high-frequency) coefficients. To reduce noise, the high-frequency coefficients (which represent fine details and noise) are selectively shrunk or set to zero:

$$W_{filtered}(x,y) = \begin{cases} W(x,y) & \text{if } |W(x,y)| \geq \tau \\ 0 & \text{if } |W(x,y)| < \tau \end{cases} \quad (2.5)$$

Where:

- $W_{filtered}(x,y)$  are the coefficients after noise reduction,
- $\tau$  is a threshold value that determines which coefficients are considered to be noise.

Finally, the inverse wavelet transform is applied to reconstruct the denoised image:

$$f_{filtred}(x,y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} W_{filtred}(x',y') \cdot \varphi(x-x',y-y') dx' dy' \quad (2.6)$$

These equations illustrate how wavelet transform can help reduce noise by selectively filtering out high-frequency noise components while preserving essential details in the image.

## 2.2.2 Contrast Enhancement

Contrast enhancement refers to image processing techniques that improve the visibility of structures within breast tissue by increasing the contrast between different tissue densities. Since medical images often have low contrast due to the similar attenuation properties of soft tissues, contrast enhancement is essential for highlighting abnormalities such as masses, microcalcifications, and architectural distortions. In Computer-Aided Diagnosis (CADx) systems, contrast enhancement is crucial in ensuring that subtle features are more distinguishable, improving the system's ability to detect potential malignancies. Standard techniques used for contrast enhancement in medical imaging include Histogram Equalization (HE), Contrast-Limited Adaptive Histogram Equalization (CLAHE), Gamma Correction, Unsharp Masking (UM), and Multi-Scale Wavelet-Based Methods [33]. These techniques can be categorized into spatial domain methods, frequency domain methods, adaptive techniques, and machine learning-based methods. Below, we explore these techniques.

### 2.2.2.1 Spatial Domain Methods

Spatial domain methods are a class of image processing techniques that operate directly on the pixel values of an image. These methods manipulate the intensity values of individual pixels or groups of pixels to enhance contrast, improve visibility of structures, or highlight specific features [34]. In CAD systems, spatial domain methods are widely used because they are computationally efficient and can be personalized to emphasize subtle details in breast tissue.

#### A. Histogram Equalization (HE)

Histogram equalization is a spatial domain technique that redistributes the intensity values of an image to improve contrast. It works by flattening and stretching the histogram of the image, making the intensity distribution more uniform [34]. HE works by transforming the original intensity histogram  $p_r(r)$  into a uniform histogram  $p_s(s)$ , using a cumulative distribution function (CDF). The transformation function is given by:

$$h(r_k) = n_k, k = 0, 1, 2, \dots, L-1 \quad (2.7)$$

where  $n_k$  is the number of pixels with intensity  $r_k$ .

### **B. Adaptive Histogram Equalization (HE)**

Adaptive Histogram Equalization (AHE) is a contrast enhancement technique that improves local contrast by applying histogram equalization to small, non-overlapping regions (tiles) of an image instead of processing the entire image globally [35]. This method enhances fine details and structures in low-contrast areas while maintaining variations in different parts of the image. After equalizing each tile, AHE uses interpolation to smooth the boundaries between adjacent regions, preventing abrupt changes in intensity. This technique is beneficial for images with varying illumination, such as mammograms, where some regions may appear darker or less contrasted than others. However, while AHE effectively enhances fine details and improves contrast in non-uniformly illuminated images, it can also significantly amplify noise, especially in homogeneous regions, making it less suitable for medical imaging without additional noise reduction techniques.

### **C. Contrast-limited Adaptive Histogram Equalization (CLAHE)**

Contrast-limited Adaptive Histogram Equalization (CLAHE) is a refined version of AHE designed to enhance local contrast in breast images without amplifying noise too much. CADx systems in medical imaging help make subtle details like microcalcifications and small masses—more visible by improving contrast in darker or less defined areas. Unlike AHE, which can exaggerate noise in uniform regions, CLAHE prevents this by setting a clip limit, which stops the histogram from over-stretching brightness levels. It then redistributes the extra intensity values to keep the image natural and smooth [36]. This makes CLAHE especially useful in breast cancer, where maintaining image clarity and reducing artifacts is crucial. However, its effectiveness depends on fine-tuning parameters like the clip limit and tile size to balance enhancing important details and keeping noise under control.

## **2.2.2.2 Frequency Domain Methods**

### **A. Fourier Transform-Based Methods**

Fourier Transform-Based Methods enhance contrast by converting an image from the spatial domain to the frequency domain, enabling selective modification of specific frequency components. In CADx systems, this approach helps highlight fine details like microcalcifications while reducing noise or unwanted variations [37]. The process begins with computing the Discrete Fourier Transform (DFT), which represents the image as a combination of sinusoidal waves, each corresponding to different frequency components. High-frequency components capture fine details and sharp edges, while low-frequency components represent smooth variations. Depending on the enhancement goal, a high-pass filter can be applied to emphasize subtle structures by suppressing low-frequency information, or a low-pass filter can smooth the image by removing high-frequency noise. Once filtering is complete, the Inverse Discrete Fourier Transform (IDFT) reconstructs the image into the spatial domain. This method provides precise control over contrast enhancement and noise suppression. However, it is computationally demanding and requires careful parameter tuning to avoid unwanted artifacts or loss of important diagnostic information.

## **B. Wavelet Transform-Based Enhancement**

Wavelet Transform-Based Enhancement is a powerful way to improve contrast in breast images by analyzing details at multiple scales. Unlike traditional methods that process the entire image, the Wavelet Transform (WT) breaks it into different layers, keeping fine details and overall structure intact. First, the image is processed with Discrete Wavelet Transform (DWT), which separates it into low-frequency components (representing smoother areas) and high-frequency components (capturing edges and tiny details like microcalcifications). This separation allows selective enhancement low frequencies can be adjusted to improve overall brightness and contrast, while high frequencies can be sharpened to highlight small but crucial abnormalities. Once the adjustments are made, the Inverse Wavelet Transform (IWT) reconstructs the image, ensuring a balanced enhancement that preserves global and local details [38]. This multi-scale approach makes WT highly effective for medical imaging, enhancing subtle structures without distorting the image. However, it is computationally demanding and requires fine-tuning to avoid excessive enhancement, which could introduce artifacts or obscure important diagnostic features.

### **2.2.3 Artifact removal**

Artifact removal is a preprocessing step in CADx systems, as undesired elements like labels, annotations, scanning artifacts, and pectoral muscle shadows can interfere with accurate analysis [39]. These artifacts can introduce noise, affect contrast adjustments, and lead to false detections, reducing the reliability of automated diagnosis. Several techniques are used to eliminate such artifacts. Thresholding and morphological operations help remove text labels and markings by distinguishing them from breast tissue based on intensity differences. Region-based segmentation methods, such as active contours (snakes) or Otsu's thresholding, are commonly applied to detect and eliminate the pectoral muscle, which can appear as a high-intensity region in the upper part of the mammogram. Deep learning-based approaches, such as convolutional neural networks (CNNs), have also been developed to automatically identify and mask artifacts, ensuring that only relevant breast tissue is analyzed. Effective artifact removal improves CADx system performance by reducing false positives, and ensuring that classification models focus on actual abnormalities rather than unnecessary elements.

### **2.2.4 Image normalization**

Image normalization is another preprocessing technique used in CADx systems for breast cancer. It helps correct variations in brightness, contrast, and intensity distribution caused by differences in imaging machines, patient anatomy, and acquisition settings [40]. By standardizing pixel intensity values, normalization ensures that mammograms are consistent and comparable, improving the accuracy of feature extraction, classification, and lesion detection. Without proper normalization, differences in image intensity can lead to misinterpretation of abnormalities and inconsistencies in CADx performance. Different approaches are used to normalize images, depending on the type of variability and the requirements of the CADx system. Different approaches are used to normalize breast images, depending on the type of variability and the requirements of the CADx system. Here are most important techniques:

### **2.2.4.1 Min-Max Normalization**

Min-Max Normalization adjusts pixel intensity values to a fixed range, such as [0,1] or [0,255], ensuring uniform brightness and contrast across mammograms. This technique preserves relative differences in contrast while preventing extreme intensity variations that could distort image interpretation [41]. It is particularly useful in CADx systems for medical imaging, where images from different imaging machines may have varying brightness levels, allowing for consistent processing, improved lesion detection, and better comparability across datasets.

### **2.2.4.2 Z-Score Normalization**

Z-Score Normalization (Standardization) adjusts pixel intensities by shifting and scaling them based on the image's average brightness and contrast (Mean and standard deviation). This means each pixel is measured relative to the overall intensity spread, ensuring the image is not affected by extreme brightness or darkness [41]. By centering the values around zero and keeping contrast differences intact, this method evens out variations while preserving important details, making images more balanced and easier to interpret.

## **2.2. Segmentation**

Segmentation is the second step in a CADx system that involves dividing a mammogram into specific areas to detect and analyze abnormalities like masses or microcalcifications accurately. A crucial part of this process is extracting the Region of Interest (ROI), which isolates suspicious areas from the surrounding breast tissue [42]. This operation ensures that only the relevant parts of the image are examined, making it easier to spot potential issues. By improving the visibility of these areas, ROI extraction allows for detailed analysis of features like shape, texture, and density, which helps differentiate between benign and malignant tumors. It also enhances the performance of classification models by providing clean, well-defined inputs for machine learning and deep learning systems. Segmentation also reduces errors, such as false positives and negatives, by concentrating on actual abnormalities rather than irrelevant background noise.

Overall, segmentation and ROI extraction streamline the diagnostic process by automating lesion detection and reducing the workload for radiologists. The process minimizes interpretation inconsistencies and makes CADx a more efficient and reliable tool for diagnosing breast cancer. Most common techniques used are thresholding-based methods, region-based segmentation, edge-based segmentation, clustering-based methods, and deep learning-based segmentation.

### **2.3.1 Thresholding-based methods**

Thresholding-based methods are image segmentation techniques that separate regions of interest (ROIs), such as potential lesions or abnormalities, from the surrounding breast tissue. These methods convert a grayscale mammogram into a binary image, where pixels are classified as either part of the ROI or the background based on their intensity values. A specific intensity threshold is set, and pixels with values above or below this threshold are assigned to different categories [43]. For example, areas with higher intensity (brighter pixels) might be

identified as suspicious masses or microcalcifications, while lower-intensity areas (darker pixels) are considered normal tissue.

Thresholding is a simple and computationally efficient approach, making it useful for initial ROI extraction. However, it may struggle with complex cases where abnormalities have similar intensity levels to surrounding tissue or when the image has uneven lighting or noise. To address these limitations, thresholding is often combined with other techniques, such as adaptive thresholding [43].

### **2.3.2 Region-based segmentation**

Region-based segmentation aims to partition an image into regions based on similarity criteria, such as intensity or texture. Methods like Region Growing and the Watershed Algorithm start from an initial seed point and iteratively expand the region by including neighboring pixels that meet predefined similarity conditions [44]. This approach efficiently segments masses with well-defined boundaries and groups pixels with shared characteristics into coherent regions. However, its performance heavily depends on the accuracy of seed point selection, and it can struggle with over-segmenting dense or heterogeneous tissues where boundaries are less distinct. Despite these limitations, region-based segmentation remains a valuable tool for isolating mammogram abnormalities.

### **2.3.3 Edge-Based Segmentation**

Edge-based segmentation is a technique that focuses on detecting and outlining the boundaries of abnormalities by identifying sharp changes in pixel intensity, known as edges. Methods like the canny edge detector and active contour (Snake) model operate by analyzing intensity gradients across the image. The canny edge detector works in multiple steps: it smoothens the image to reduce noise, then computes the gradient magnitude and direction to highlight areas of rapid intensity change. Non-maximum suppression is applied to thin the edges, and finally, hysteresis thresholding is used to distinguish evident edges from blurred ones, ensuring only significant boundaries are retained. On the other hand, the active Contour model starts with an initial curve (often drawn near the lesion) and iteratively deforms it to fit the lesion's boundaries [45]. This deformation is driven by energy minimization, where the curve adjusts itself to align with edges while maintaining smoothness. Internal forces keep the contour cohesive, while external forces pull it toward edges based on gradient information.

These methods are particularly effective for segmenting lesions with well-defined edges, as they rely on clear intensity transitions. However, they can struggle in noisy images or when boundaries are faint, as weak gradients may lead to incomplete or inaccurate edge detection. Despite these limitations, edge-based segmentation remains a powerful tool for outlining mammogram abnormalities with distinct boundaries.

### **2.3.4 Clustering-based segmentation**

Clustering-based segmentation methods are techniques that group pixels into clusters based on shared features such as intensity, texture, and sometimes even higher-level statistical information. In methods like K-Means Clustering, the algorithm begins by selecting initial cluster centroids, often randomly, and then iteratively assigns each pixel to the nearest centroid

based on a chosen distance metric (usually Euclidean distance), recalculating the centroids as the mean of the assigned pixels until convergence is achieved [46]. Fuzzy C-Means (FCM) extends this approach by allowing pixels to belong to multiple clusters with varying degrees of membership, which helps capture ambiguous boundaries and offers a smoother transition between regions. These methods work by leveraging the intrinsic similarities within local pixel neighborhoods, effectively partitioning the image into regions that correspond to different tissue types or abnormalities, such as masses with varying intensities in mammograms [46]. However, they require the predefined number of clusters, which can be challenging to set optimally, and they are sensitive to noise necessitating preprocessing steps like smoothing or filtering to prevent erroneous cluster assignments and ensure robust segmentation in complex, real-world imaging scenarios.

### **2.3.5 Deep learning-based segmentation**

Deep learning-based segmentation use convolutional neural networks (CNNs) to automatically learn and extract multi-level features from breast images for precise lesion delineation. Architectures such as U-Net and Mask R-CNN embody this approach: U-Net employs an encoder-decoder structure with skip connections that fuse high-resolution features from the contracting path with the spatial details recovered during expansion, ensuring that both global context and fine-grained information are retained. Mask R-CNN, on the other hand, extends the Faster R-CNN framework by incorporating an additional branch dedicated to generating pixel-level masks for each proposed region of interest, effectively coupling object detection with detailed segmentation [47]. These models are typically trained end-to-end on large, annotated datasets using loss functions like cross-entropy or Dice coefficient loss to directly optimize segmentation accuracy. Additionally, techniques such as data augmentation and transfer learning are commonly applied to enhance model robustness and generalization, especially in the context of variable image quality and limited data availability. Despite their high accuracy and adaptability, these networks require significant computational resources and extensive, high-quality labeled data, posing challenges for widespread clinical implementation.

## **2.4 Features extraction**

Feature extraction in Computer-Aided Diagnosis (CADx) systems is the third step and a fundamental process that transforms raw mammogram images into a set of quantitative attributes known as features that capture critical characteristics of breast tissue and potential lesions. By converting high-dimensional image data into a compact, informative representation, feature extraction enables automated algorithms to effectively analyze and classify lesions, and improving diagnostic accuracy and reliability.

### **2.4.1 Features**

Features (or attributes) refer to quantifiable properties extracted from medical images that capture essential details about the breast tissue and potential lesions. These features can include measurements of intensity, texture, shape, and morphology, and they serve as inputs to classifiers (machine learning or deep learning algorithms). Essentially, they provide a compact and informative representation of the raw image data, calculating measurable properties that

help distinguish between normal tissue and abnormalities such as masses or microcalcifications in classification.

Features can be broadly categorized into several types:

- **Intensity Features:** These include basic pixel values and statistical measures such as the mean, variance, skewness, and kurtosis of pixel intensity distributions. They provide insights into tissue density and can highlight areas of abnormal brightness or darkness associated with lesions.
- **Texture Features:** Texture describes the spatial arrangement of pixel intensities. Techniques like Gray-Level Co-occurrence Matrices (GLCM) and Local Binary Patterns (LBP) are used to quantify texture by capturing the frequency and pattern of pixel intensity variations [48,49]. Gabor filters further enhance this by analyzing frequency and orientation, allowing for the detection of fine structural details within the tissue.
- **Shape and Morphological Features:** These describe the geometric properties of regions of interest (ROIs), such as lesions [48,49]. Shape descriptors include metrics like area, perimeter, and compactness, while more advanced descriptors such as moment invariants, Fourier descriptors, and Zernike descriptors quantify contours, symmetry, and boundary irregularities. These descriptors are designed to be invariant to changes in scale, rotation, and translation, making them robust for comparing lesions across different images.
- **Automatically Learned Features:** are obtained using deep convolutional neural networks (CNNs) that learn hierarchical representations directly from raw images. Initially, the CNN extracts low-level features such as edges and textures using convolutional filters; as the network deepens, it aggregates these into higher-level, more abstract representations like shapes and lesion boundaries. The learning process involves backpropagation, gradient descent, and non-linear activations (e.g., ReLU), along with pooling to reduce spatial dimensions and enhance invariance. These multi-scale, high-dimensional features are optimized end-to-end and are particularly effective in capturing complex patterns, ultimately improving diagnostic accuracy by adapting to variations in lesion appearance and image quality [50].

### 2.4.2 Feature selection

Feature selection is the process of identifying and choosing a subset of the most relevant features from an initial set of extracted attributes, thereby reducing the dimensionality of the dataset used in the CADx system. The primary goal is to improve model performance by eliminating redundant or irrelevant features, which can lead to overfitting and increased computational complexity. By focusing on the most discriminative features, feature selection improves the accuracy, efficiency, and interpretability of the diagnostic models.

Several techniques are used for feature selection, generally classified into three categories:

- **Filter Methods:** These techniques assess the relevance of features based on statistical measures, independently of the learning algorithm. Examples include correlation coefficients, chi-square tests, mutual information, and variance thresholds [50,51]. Filter

methods are computationally efficient and provide a fast way to eliminate irrelevant features.

- **Wrapper Methods:** Wrapper methods evaluate feature subsets by directly measuring the performance of a specific predictive model. Techniques such as recursive feature elimination (RFE) and sequential feature selection fall under this category [50,51]. Although they often provide better performance by considering feature interactions, they are typically more computationally intensive.
- **Embedded Methods:** Embedded techniques incorporate feature selection as part of the model training process. Methods like LASSO (L1 regularization) and decision tree-based algorithms (e.g., random forests) automatically select features by assigning importance weights during model fitting [49,50,51]. These approaches strike a balance between computational efficiency and performance by integrating the selection process within model optimization.

In the context of breast cancer, effective feature selection is crucial for handling the high-dimensional data that results from extracting statistical, texture, shape, and automatically learned features. By reducing the feature space, the CADx system can focus on the most clinically significant attributes, which increases diagnostic accuracy, reduces false positives and negatives, and rationalizes the computational workload.

## 2.5 Classification

Classification is the final stage in a computer-aided diagnosis (CADx) system. It is the process of assigning discrete labels to input data based on learned patterns, typically following the extraction and selection of discriminative features, which is based on the segmentation result, to perform automatic diagnosis. In practice, classifiers are trained on labeled datasets where each mammogram is associated with a ground-truth diagnosis. Algorithms such as Support Vector Machines (SVMs), Random Forests, Logistic Regression, and deep neural networks (DNNs) are commonly employed due to their ability to model complex, non-linear relationships in feature space.

Classification in machine learning includes several paradigms and types, each suited to different data structures and objectives. The primary paradigms include supervised learning, where models are trained on labeled data to predict outputs; unsupervised classification (clustering), which identifies patterns or clusters in unlabeled data; and reinforcement learning, where agents learn optimal actions through trial-and-error interactions with an environment to maximize cumulative rewards. Within classification tasks, common types include binary classification (distinguishing between two classes, e.g., malignant vs. benign), multi-class classification (assigning one label among three or more mutually exclusive classes, e.g., BI-RADS scores 1–6), and multi-label classification applies when multiple abnormalities (e.g., masses, calcifications) coexist in a single scan, requiring simultaneous detection. Additional variations include imbalanced classification (handling skewed class distributions) and hierarchical classification (organizing labels into nested structures).

### **2.5.1 Supervised Classification**

Supervised classification in CADx breast cancer involves training algorithms on labeled datasets where each mammogram is annotated with the ground truth (e.g., benign or malignant). In this paradigm, classifiers learn to map high-dimensional feature vectors comprising statistical, textural, morphological, and automatically learned features to diagnostic outcomes. Techniques commonly employed include Support Vector Machines (SVMs), Random Forests, Logistic Regression, and deep learning architectures like Convolutional Neural Networks (CNNs) [52,53]. During training, these models optimize objective functions (e.g., cross-entropy or hinge loss) using iterative methods such as gradient descent and employ regularization techniques to prevent overfitting.

Supervised methods in medical imaging rely on extensive datasets of images that have been meticulously annotated by experts, such as radiologists, with definitive labels indicating benign or malignant lesions. This abundance of high-quality, labeled data is crucial because it allows machine learning models to learn complex mappings between extracted features and clinical outcomes. The presence of ground-truth labels not only facilitates the training process but also ensures that the models can generalize well to new, unseen data.

In addition, supervised methods provide clear, interpretable performance metrics that are vital in clinical settings. For example, accuracy, sensitivity (or recall), specificity and many others metrics that will be detailed in this chapter. These metrics offer a comprehensive evaluation of a model's diagnostic efficacy, ensuring that the CADx system meets the stringent reliability and safety standards required in clinical practice.

### **2.5.2 Unsupervised Classification**

Unsupervised classification, often referred to as clustering, does not rely on labeled data but instead seeks to identify inherent patterns or groupings within the feature space. In the context of CADx for breast cancer, unsupervised methods can be used to discover latent structures or to segment mammograms into regions with similar characteristics, which may correspond to different tissue types or potential anomalies. Techniques such as K-means clustering, hierarchical clustering, and self-organizing maps (SOM) are utilized, relying on distance metrics (e.g., Euclidean distance) to group similar feature vectors. While unsupervised learning does not directly produce diagnostic labels, it can assist in anomaly detection or pre-processing steps by highlighting clusters of images that deviate from typical patterns [52,53]. These clusters can later be interpreted in conjunction with clinical expertise or used to guide semi-supervised learning processes.

### **2.5.3 Reinforcement Learning**

Reinforcement learning (RL) represents an emerging paradigm in CADx systems, where an agent learns to make sequential decisions by interacting with an environment to maximize a cumulative reward. In breast cancer, RL can be employed to optimize the classification process, such as dynamically selecting regions of interest (ROI) for further analysis or refining segmentation boundaries in a sequential manner. Techniques in this domain include Q-learning, Deep Q-Networks (DQN), and policy gradient methods [52,53]. The RL agent receives feedback in the form of rewards derived from metrics such as diagnostic accuracy or lesion

detection performance which guide it in learning an optimal policy for decision making. For example, an RL-based system might learn to adjust its focus within a mammogram based on the likelihood of detecting subtle abnormalities, permitting the improvement of both efficiency and accuracy. This approach is particularly useful in active learning settings, where the system iteratively refines its decision-making strategy based on real-time performance, eventually adapting to variations in image quality and lesion presentation.

## **2.5.4 Challenges of classification**

Classification generally and in the context of CADx systems for breast cancer specially, models face several challenges that can significantly impact diagnostic performance.

### **2.5.4.1 Overfitting**

Occurs when a model learns patterns that are highly specific to the training dataset, including noise and outliers, rather than capturing the underlying generalizable relationships. In medical imaging, this risk is aggravated by the high-dimensionality of extracted features such as textural, morphological, and automatically learned attributes and the relatively limited size of annotated datasets. As a result, a model might achieve excellent performance on the training data but perform poorly on unseen images, leading to unreliable diagnostic predictions. To mitigate overfitting, various techniques are employed, including regularization methods (e.g., L1, L2 penalties, dropout), cross-validation strategies, and early stopping during training. Additionally, simplifying the model architecture or employing dimensionality reduction techniques can help ensure that the classifier generalizes well to new patient data [54].

### **2.5.4.2 Imbalanced Data**

Imbalanced Data is another significant challenge in CADx. In many clinical datasets, the distribution of classes is unequal, with a disproportionately high number of benign cases compared to malignant ones. This class imbalance can bias the model toward predicting the majority class, thereby compromising its sensitivity to the minority class often the more clinically critical (malignant cases). The consequences of such bias include high overall accuracy but low recall (sensitivity) for detecting malignant lesions, which is unacceptable in a clinical setting.

To address this issue, various strategies are implemented, such as resampling techniques (oversampling the minority class or undersampling the majority class), synthetic data generation methods like SMOTE (Synthetic Minority Over-sampling Technique), and the use of cost-sensitive learning where misclassification penalties are adjusted based on the class distribution [54,55]. Moreover, employing evaluation metrics beyond simple accuracy such as precision, recall, F1-score, and the area under the ROC curve (AUC) provides a more nuanced assessment of the classifier's performance in the presence of imbalanced data. Despite the availability of large datasets, the number of malignant cases is often relatively low compared to benign cases. This scarcity, especially for complex or rare lesion types, intensifies class imbalance and may interfere with the classifier's ability to learn robust patterns for minority classes.

### **2.5.4.3 Computational complexity**

Deep learning models arise primarily from the high number of trainable parameters that these networks possess. Modern architectures, especially those designed for image analysis like convolutional neural networks (CNNs), often consist of millions of parameters distributed across numerous layers. This high parameter count enables the network to learn detailed and nuanced features from the data; however, it also significantly increases the computational resources required for both training and inference [55].

During training, the optimization process involves numerous matrix multiplications, convolutions, and backpropagation steps across all layers, which demands extensive processing power and memory. These operations are typically executed on specialized hardware, such as GPUs or TPUs, to accelerate computation. The reliance on such hardware not only increases the operational costs but can also introduce challenges when deploying the system in real-time clinical environments where immediate processing and decision-making are critical.

Inference, or the process of making predictions using a trained model, also suffers from the model's complexity. High parameter counts translate to increased latency, as the model needs to perform a significant number of computations to arrive at a diagnosis. This can be particularly problematic in scenarios that require real-time analysis, such as during screening sessions or intraoperative decision support [54].

Furthermore, the computational demands necessitate careful model design and optimization strategies. Techniques such as model pruning, quantization, and knowledge distillation are often employed to reduce the number of parameters and computational overhead while attempting to preserve diagnostic accuracy [55]. Balancing the trade-off between model complexity and computational efficiency remains a key challenge in the practical deployment of deep learning-based CADx systems in clinical settings, as reducing latency and cost without compromising performance is essential for widespread adoption [56].

#### **2.5.4.4 Interpretability of Complex Models**

Advanced models, especially deep neural networks, despite they deliver highly accurate classification, are often perceived as "black boxes", making it difficult to interpret their decision-making process. For deep learning to gain clinical acceptance, the algorithms must be interpretable and explainable. Interpretability, or transparency, refers to the degree to which a human observer can understand the model's reasoning process. Explainability, on the other hand, refer to the fundamental attributes of a model that describe its internal mechanisms and help clarify or "explain" its decisions. In clinical settings, the lack of transparency can be a barrier to trust and acceptance by practitioners who require explainable and justifiable diagnostic conclusions [56,57].

## **2.6 Metrics and evaluation of CADx systems**

Evaluation of Computer-Aided Diagnosis (CADx) systems for breast cancer and medical imaging in general, relies on a comprehensive set of metrics to assess diagnostic performance, robustness, and clinical utility. These metrics can be categorized into classification metrics, segmentation metrics, and evaluation strategies.

### **2.6.1 Classification Metrics**

In the context of lesion classification (e.g., benign vs. malignant), key performance indicators include:

- **Confusion matrix:** is a fundamental tool in evaluating the performance of classification models, especially in the context of CADx systems [58]. It is a tabular representation that summarizes the number of correct and incorrect predictions made by the model, organized by the actual and predicted classes as show in Table 2.1. In a binary classification scenario where the task is to distinguish between malignant and benign lesions the confusion matrix consists of four key components:
  - **True Positives (TP):** The number of malignant lesions correctly identified as malignant.
  - **False Positives (FP):** The number of benign lesions incorrectly classified as malignant.
  - **True Negatives (TN):** The number of benign lesions correctly identified as benign.
  - **False Negatives (FN):** The number of malignant lesions incorrectly classified as benign.

Predicted Labeled	Predicted Positive (Malignant)	Predicted Negative (Benign)
Actual Positive (Malignant)	True Positives (TP)	False Negatives (FN)
Actual Negative (Benign)	False Positives (FP)	True Negatives (TN)

**Table 2.1 :** Confusion matrix table.

- **Accuracy:** The proportion of total correct predictions (both benign and malignant) to the overall number of cases. While simple to interpret, accuracy alone may be misleading in datasets with imbalanced classes.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.8)$$

- **Sensitivity (Recall):** Also known as the true positive rate, sensitivity measures the proportion of actual malignant cases correctly identified by the system. This metric is crucial in a clinical setting were failing to detect a malignant lesion can have severe consequences.

$$Sensitivity = \frac{TP}{TP+FN} \quad (2.9)$$

- **Specificity:** The true negative rate that quantifies the proportion of benign cases correctly identified. High specificity reduces the number of false positives, thereby minimizing unnecessary biopsies and associated patient anxiety.

$$Specificity = \frac{TN}{TN+FP} \quad (2.10)$$

- **Precision (Positive Predictive Value):** This metric reflects the proportion of true malignant predictions among all cases classified as malignant. It is especially important in evaluating the clinical relevance of detected lesions.

$$Precision = \frac{TP}{TP+FP} \quad (2.11)$$

- **F1-Score:** The harmonic means of precision and recall, providing a balanced measure that is particularly informative when dealing with imbalanced datasets.

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} = 2 \times \frac{TP}{2 \times TP+FP+FN} \quad (2.11)$$

- **Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC):** The ROC curve plots sensitivity versus 1-specificity across various threshold settings, and the AUC provides a single scalar value summarizing the model's ability to discriminate between classes over all thresholds. A high AUC indicates robust performance across different operating points [58].

## 2.6.2 Segmentation Metrics

For CADx systems that include lesion segmentation, evaluation metrics focus on the spatial accuracy of delineating regions of interest [59]:

- **Dice Coefficient (F1 Score for Segmentation):** Measures the overlap between the segmented lesion and the ground truth, defined as twice the area of overlap divided by the total number of pixels in both the predicted and ground truth masks. It is sensitive to both false positives and false negatives.
- **Jaccard Index (Intersection over Union, IoU):** Similar to the Dice coefficient, it quantifies the similarity between the predicted segmentation and the reference standard by dividing the intersection of the predicted and true regions by their union.
- **Hausdorff Distance:** Assesses the maximum distance of the predicted segmentation boundary to the ground truth boundary, providing insight into the worst-case segmentation error.

## 2.6.3 Evaluation Strategies

Beyond individual metrics, robust evaluation of CADx systems is achieved through methodological strategies that ensure generalizability and reliability [60]:

- **Cross-Validation:** Techniques such as k-fold cross-validation help assess model performance across multiple subsets of data, decreasing the risk of overfitting and ensuring that performance metrics are not artifacts of a particular data split.
- **Independent Test Sets:** Utilizing an independent, often external, dataset to evaluate the CADx system provides an unbiased assessment of its generalizability across different patient populations and imaging conditions.
- **Statistical Analysis:** Confidence intervals, p-values, and other statistical tests are often employed to determine the significance of observed performance differences,

particularly when comparing multiple models or assessing improvements over baseline methods.

- **Clinical Validation:** Finally, CADx systems must be evaluated in real-world clinical settings. Prospective studies and reader studies, where radiologists interact with the system, provide essential feedback on usability, diagnostic accuracy, and potential integration into clinical workflows.

In summary, the evaluation of CADx systems for breast cancer covers a variety of performance metrics and methodological approaches to ensure that the system is both accurate and clinically viable. By combining classification metrics, segmentation evaluation, and rigorous validation strategies, researchers and clinicians can effectively measure the system's potential impact on patient results and its readiness for clinical deployment.

### 3. Conclusion

In conclusion, the development and implementation of Computer-Aided Diagnosis (CADx) systems for automatic diagnosis of breast cancer have emerged as transformative approaches in the early detection and diagnosis of breast cancer. The chapter has comprehensively discussed the critical steps involved in these systems from segmentation and feature extraction to classification and performance evaluation, while also highlighting the significant challenges that must be overcome to ensure robust clinical application. Based on the comprehensive review presented in this chapter, I have chosen to develop a shape descriptor that leverages traditional machine learning techniques with robust classifiers like SVM and Random Forest; alongside an alternative deep learning approach for CADx in breast cancer based on convolutional neural networks (CNNs) to perform accurate automatic diagnosis of masses and microcalcifications.

# Chapter 3: Image description model for breast cancer recognition

## 1. Introduction

The accurate description and analysis of breast lesions in breast cancer imaging are essential for early breast cancer detection and diagnosis. Mammography remains one of the most effective screening techniques, enabling the identification of abnormalities that may indicate the presence of malignant tumors [95]. However, the classification of these abnormalities depends on the ability to extract and interpret relevant features that characterize lesions [61]. To achieve this, descriptors play a fundamental role in translating visual and pathological properties of lesions into mathematical representations. These descriptors serve as critical elements in computer-aided diagnosis (CADx) systems, helping to improve classification accuracy and assist radiologists in making informed decisions.

A particular challenge in breast cancer diagnosis is the detection and classification of spiculated masses, which are strongly associated with malignancy [62]. These masses exhibit irregular, radiating spicules that make their boundary difficult to define, increasing the complexity of automated lesion characterization. Despite their clinical significance, few studies have specifically addressed the development of descriptors tailored for spiculated masses.

To address the challenge of analyzing spiculated masses, the second part of this chapter will be dedicated to the newly proposed descriptor: Polygon Approximation Triangle-Area Representation (PATAR). This descriptor is designed to better capture the unique boundary and structural characteristics of spiculated lesions, enhancing their differentiation from other types of breast masses. By applying polygon approximation, PATAR effectively extracts concave and convex spaces along the lesion boundary, key indicators of shape irregularities commonly associated with malignancies.

These concave and convex regions are quantified using Triangle-Area Representation (TAR) signature allowing to this method to well estimate the degree of malignancy of breast mass. By integrating key aspects of shape and contour irregularity, our proposed method aims to improve classification accuracy and provide a more reliable tool for computer-aided diagnosis (CADx). The first section of this chapter will be dedicated to a presentation of descriptors in literature; shape and geometric descriptors founded on basic measurement, and texture descriptors. Also in this chapter, we will present the methodology behind PATAR, its mathematical formulation, and its evaluation against existing shape descriptors, demonstrating its potential in improving breast cancer detection.

## 2. Shape descriptors (Morphological)

### 2.1 Geometric descriptors

Geometric descriptors are quantitative measurements that characterize the morphological properties of a region of interest (ROI) in a medical image. In mammography, these descriptors capture the geometric attributes like compactness, area, rectangularity of breast masses, which are critical since malignant lesions often have distinctive shape characteristics compared to benign ones. In the sections below, some of the most commonly used shape descriptors will be detailed.

#### 2.1.1 Center of Gravity (Centroid)

The centroid of a shape is a geometric property used in various shape descriptors for mammographic image analysis. It represents the center of mass or geometric center of a lesion and serves as a reference point for analyzing boundary irregularities, asymmetry, and shape complexity. In Computer-Aided Diagnosis systems, centroid-based descriptors are particularly useful for classifying breast lesions, detecting spiculated masses, and evaluating shape distortions associated with malignancies [63].

The centroid of a shape is the average position of all points that define the shape's boundary or region. Mathematically, it is given by:

$$C_x = \frac{1}{N} \sum_{i=1}^N x_i \quad (3.1)$$

$$C_y = \frac{1}{N} \sum_{i=1}^N y_i \quad (3.2)$$

- $C_x, C_y$  are the centroid coordinates,
- $x_i, y_i$  are the pixel coordinates of the lesion,
- $N$  is the number of pixels defining the lesion.

#### 2.1.2 Perimeter

Is the total length of the boundary of a mass measured by summing the distances between consecutive boundary pixels, providing a fundamental measurement used in calculating various shape ratios and serving as a basic indicator of lesion size that helps differentiate between smooth-bordered benign masses and irregular malignant ones [63,66].

#### 2.1.3 Area

Represents the total number of pixels contained within the mass boundary, calculated by simple counting of pixels in the segmented region (ROI), providing a basic size measurement that helps assess the extent of the lesion and may indicate growth when monitored across sequential mammograms [64].

### 2.1.4 Circularity

The circularity descriptor is a shape-based feature used in Computer-Aided Diagnosis (CADx) systems for mammography to assess lesion regularity [64,65]. It measures how closely a shape resembles a perfect circle, with benign masses typically having higher circularity (smooth, well-defined edges) and malignant tumors, especially spiculated masses, exhibiting lower circularity due to irregular, jagged, or radiating contours. Circularity is calculated using the formula:

$$C = \frac{4 \pi \cdot Area}{Perimeter^2} \quad (3.3)$$

Where:

- *Area* is the lesion area,
- *Perimeter* is its perimeter.

A perfect circle has  $C = 1$ , while irregular shapes have  $C < 1$ . In CADx mammography, circularity is used alongside other shape descriptors (e.g., convexity, compactness) to differentiate benign from malignant lesions, improving breast cancer detection and classification accuracy.

### 2.1.5 Rectangularity

Rectangularity (also called extent or rectangular filling ratio) is a shape descriptor used to measure how well a given shape (such as a breast lesion in a mammogram) fits within its minimum bounding rectangle. It quantifies how efficiently a lesion occupies its surrounding rectangular region, providing insights into shape regularity and compactness [65,67].

Rectangularity is defined as the ratio of the lesion's area to the area of its minimum bounding rectangle:

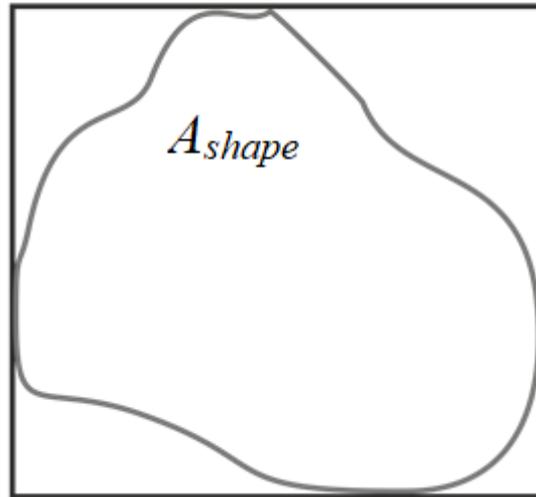
$$R = \frac{A_{shape}}{A_{bounding\ rectangle}} \quad (3.4)$$

Where :

- $R$  = Rectangularity (ranges from 0 to 1, with 1 indicating a perfect rectangle).
- $A_{shape}$  = Area of the lesion (number of pixels inside the lesion).

- $A_{\text{bounding rectangle}}$  = Area of the minimum bounding rectangle (width  $\times$  height of the smallest rectangle enclosing the lesion).

*A<sub>bounding rectangle</sub>*



**Figure 3. 1:** Rectangularity measure.

However, this rectangularity formula does not consider the orientation of the object. A perfect, upright rectangle has a rectangularity value of 1, but when the same rectangle is tilted, its rectangularity value changes. This makes the descriptor sensitive to rotation.

### 2.1.6 Compactness

The compactness descriptor is a shape-based feature used in Computer-Aided Diagnosis (CADx) systems for mammography to assess how efficiently a lesion occupies its space. It helps differentiate smooth, well-defined benign masses from irregular, invasive malignant tumors. Compactness is calculated using the formula:

$$C = \frac{P^2}{A} \quad (3.5)$$

Where  $P$  is the lesion's perimeter and  $A$  is its area. Lower compactness values indicate more regular, compact shapes (e.g., benign cysts), whereas higher compactness values suggest irregular, complex contours (e.g., spiculated malignant tumors). Since malignant lesions often have higher perimeter-to-area ratios due to their rough, invasive growth, compactness serves as a key parameter in CADx mammography [65,66].

### 2.1.7 Fourier descriptor

The Fourier Descriptor (FD) is a shape-based feature used in Computer-Aided Diagnosis (CADx) systems for mammography to analyze lesion contour irregularities by transforming the

lesion's boundary into the frequency domain using the Discrete Fourier Transform (DFT). The lesion contour is first extracted and represented as a sequence of complex numbers  $z_n = x_n + jy_n$ , then transformed using DFT:

$$F(k) = \sum_{n=0}^{N-1} z_n e^{-j2\pi kn/N} \quad (3.6)$$

Where low-frequency components capture global shape features (e.g., circularity of benign masses) and high-frequency components highlight local irregularities (e.g., spiculations in malignant tumors). To ensure invariance to rotation, scale, and translation, the Fourier coefficients are normalized by setting  $F(0)$  to zero and scaling by  $|F(1)|$ . In mammographic CADx applications, benign lesions (cysts, fibroadenomas) typically have dominant low-frequency components, indicating smooth, well-defined borders, while malignant tumors (spiculated masses, IDC) show higher-frequency variations, reflecting irregular, invasive growth. The Fourier Descriptor is particularly valuable for detecting architectural distortions, making it a powerful tool in breast cancer detection and classification [68].

## 2.2 High order descriptors

High-order shape descriptors are advanced techniques used to characterize the complex geometric and morphological properties of regions of interest (ROIs), such as masses, microcalcifications, or architectural distortions. These descriptors go beyond basic shape metrics (e.g., area, perimeter) to capture and explore intricate details about the structure, boundaries, and spatial relationships within the image. Below are examples of high-order shape descriptors used in mammography.

### 2.2.1 Moments based descriptors

Moment-based shape descriptors are mathematical tools used to characterize the shape of objects in images by computing moments, which are weighted averages of pixel intensities. These descriptors are widely used in mammography to analyze the shape and structure of lesions, such as masses and microcalcifications. Below are the main types of moment-based shape descriptors, each explained in detail:

#### 1- Geometric moments

Geometric moments are the most basic type of moments, calculated as weighted sums of pixel intensities over the image. The moment of order  $(p + q)$  is given by;

$$m_{pq} = \sum_x \sum_y x^p y^q I(x, y) \quad (3.7)$$

Where  $I(x, y)$  is the intensity at pixel  $(x, y)$ .

These moments capture global shape properties, such as the centroid, orientation, and size of an object. However, they are not invariant to transformations like rotation or scaling, making them less robust for shape analysis in mammography without additional normalization [69,70].

## 2- Hu Invariant Moments (Invariant)

Hu moments are derived from geometric moments and provide shape descriptors that are invariant to translation, rotation, and scaling. This means that a lesion's shape can be identified regardless of imaging angle, size, or position. Hu defined seven invariant moments using normalized central moments  $\eta_{pq}$ , where:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{(1+\frac{p+q}{2})}} \quad (3.8)$$

with

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x, y) \quad (3.9)$$

where  $\bar{x}, \bar{y}$  are the centroid coordinates of the shape [62].

## 3- Zernike Moments (ZMs)

Zernike moments are computed using a set of orthogonal polynomials defined over a unit circle, making them invariant to rotation and robust against noise. They are particularly useful for analyzing lesions with complex boundary structures.

The Zernike moment of order  $\bar{n}$  and repetition  $\bar{m}$  is defined as

$$Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y I(x, y) V_{nm}^*(x, y) \quad (3.10)$$

where  $V_{nm}(x,y)$  is the Zernike polynomial:

$$V_{nm}(r, \theta) = R_{nm}(r) e^{jm\theta} \quad (3.11)$$

with  $R_{nm}(r)$  being the radial polynomial.

In addition to Geometric, Zernike and Hu moments, other moment-based shape descriptors, such as Legendre Moments (LMs) and Wavelet Moments (WMs), are widely used in mammographic analysis. Legendre Moments are computed using Legendre polynomials, which are orthogonal over a rectangular domain, making them efficient for analyzing lesions in image regions with well-defined boundaries. They provide a compact representation of shape and are particularly useful for detecting lobulated or irregular tumor contours. On the other hand, Wavelet Moments integrate both spatial and frequency-domain information, allowing multi-scale shape analysis [69,71]. These moments are derived from wavelet transforms, which decompose an image into different resolution levels, capturing both global shape characteristics and fine structural details such as microcalcifications or spiculations in malignant tumors. The combination of these moment-based descriptors enhances feature extraction in CADx systems,

improving the accuracy of breast cancer detection by distinguishing between smooth benign lesions and complex malignant masses.

### 2.2.2 Fractal analysis

Fractal analysis is a powerful mathematical approach used in mammography to quantify the complexity and irregularity of breast lesion shapes. Unlike traditional shape descriptors that rely on smooth boundaries, fractal analysis is particularly effective for characterizing highly irregular, spiculated, or infiltrative tumor margins, which are common in malignant breast lesions [64].

The fractal dimension (FD) is the primary measure used in fractal analysis to quantify shape complexity. It describes how a shape's detail changes with scale, providing insights into the lesion's structural irregularity. A higher fractal dimension indicates greater complexity and roughness, which is often associated with malignancy, whereas lower values suggest smooth, well-defined benign lesions. One of the most widely used methods to compute FD in medical imaging is the box-counting method, defined as:

$$D = \lim_{r \rightarrow 0} \frac{\log N(r)}{\log(1/r)} \quad (3.12)$$

where:

- $N(r)$  is the number of boxes of size  $r$  required to cover the shape boundary,
- $D$  is the estimated fractal dimension.

This method applies a grid of different-sized squares to an image and counts how many squares contain part of the lesion boundary. As the grid becomes finer, the fractal dimension is estimated based on how the count changes with scale [73,74].

Fractal analysis in medical imaging is particularly useful for distinguishing between benign and malignant lesions based on boundary complexity. Malignant lesions typically show higher fractal dimensions (e.g.,  $D > 1.3$ ) due to their irregular, spiculated, or infiltrative growth patterns, whereas benign lesions tend to have lower fractal dimensions (e.g.,  $D < 1.2$ ), reflecting their smooth and well-defined contours. One of its key advantages is scale invariance, allowing the detection of lesion complexity across multiple scales, making it robust against image resolution differences.

### 2.2.3 Skeletonization-Based Descriptors

Skeletonization-Based Descriptors are a class of shape analysis tools that focus on extracting and analyzing the medial axis or skeleton of an object. The skeleton is a simplified, thin-line representation of the shape that preserves its topological and geometric properties. These descriptors are particularly useful for analyzing complex shapes, such as spiculated

masses or architectural distortions in mammography. Below is a detailed explanation of skeletonization-based descriptors, including their definition, functioning, and applications [75].

Skeletonization is a process that reduces a 2D shape to its medial axis or skeleton, which is a set of curves or lines that lie along the center of the shape. The skeleton captures the essential structure of the shape, including its branches, endpoints, and loops, while removing redundant information about its thickness. Skeletonization-based descriptors are derived from this skeleton and are used to quantify its properties, such as branch length, curvature, and connectivity.

The skeleton is typically extracted using thinning algorithms, such as Zhang-Suen or Morphological Thinning, which iteratively remove boundary pixels while preserving the object's connectivity and topology. The result is a 1-pixel-wide representation of the shape's medial axis.

Once the skeleton is obtained, various features are computed to describe its properties:

- **Branch Points:** Points where the skeleton splits into multiple branches.
- **End Points:** Terminal points of the skeleton.
- **Branch Length:** The length of individual branches in the skeleton.
- **Curvature:** The degree of bending or curvature along the skeleton.
- **Loop Detection:** Identification of closed loops in the skeleton.
- **Symmetry Analysis:** Measures of symmetry or asymmetry in the skeleton.

These features are quantified to create a set of descriptors that capture the shape's complexity, topology, and geometry [75].

Skeletonization-based descriptors offer several advantages, including topological preservation, which ensures that the skeleton maintains the object's connectivity and structure, making it efficient for analyzing complex shapes like spiculated masses in mammography. Additionally, they provide dimensionality reduction by simplifying the shape to its essential medial axis, and reducing computational complexity. These descriptors are also robust to variations in shape thickness or boundary irregularities, making them reliable for characterizing lesions. However, they face challenges such as sensitivity to noise, which can introduce false branches, computational complexity due to the intensive processing required for large or intricate shapes, and parameter sensitivity, as the quality of the skeleton depends heavily on the choice of thinning algorithm and parameters. Despite these challenges, skeletonization remains a powerful tool for shape analysis in medical imaging [76,77].

### 3. Texture descriptors

Texture is a fundamental visual attribute used in computer vision to analyze surface patterns and distinguish objects based on their appearance. Texture description involves extracting numerical features that characterize the spatial distribution of pixel intensities in an image [78]. These descriptors help in various applications, such as medical imaging. Texture descriptors are generally categorized into four primary types: structural, statistical, model-based, and transform methods. Each category represents a different approach to analyzing texture, ranging from pattern-based representations to mathematical transformations.

#### 3.1 Structural methods

In structural methods, texture is defined by the fundamental units and their spatial arrangement. These fundamental units or primitives can be as simple as individual pixels, regions, or line-like shapes. The spatial organization of these units is determined by analyzing their geometric relationships or statistical properties. Essentially, structural approaches treat texture as a composition of basic patterns. Once these primary patterns are identified, their statistical properties are computed and used as features. These methods are well-suited for textures with regular and repetitive structures. However, they are less effective for images with irregular or complex textures, as they may struggle to capture the inherent variability and randomness in such cases [78].

##### 3.1.1 Structural Element Analysis (Primitive-Based Methods)

This method breaks a texture into fundamental primitives (e.g., lines, edges, shapes) and analyzes their spatial distribution. A key example is Edge Frequency Analysis, which detects edge primitives and their orientations to define texture patterns. The Edge Density Calculation is given by:

$$D_E = \frac{\sum_{ij} |\nabla f(i, j)|}{N} \quad (3.13)$$

where  $\nabla f(i, j)$  represents the gradient magnitude at pixel  $(i, j)$ , and  $N$  is the total number of pixels. In mammography, this method can be used for detecting spiculated masses (characterized by radiating edges) and identifying sharp transitions in malignant tumors. Its strengths include effectiveness in highlighting lesion boundaries and detecting architectural distortions. However, it is sensitive to image resolution and noise, and edge-based analysis alone may fail to capture fine-grained texture variations in dense breast tissue [79].

##### 3.1.2 Morphological Analysis (Mathematical Morphology)

This a structural method analyzes texture by applying morphological operations (e.g., dilation, erosion, opening, and closing) to extract shape-based features. A key example is Morphological Closing, which enhances mass detection by connecting fragmented structures. The operation is defined as:

$$A \bullet B = (A \oplus B) \ominus B \quad (3.14)$$

where  $A$  is the binary image,  $B$  is the structuring element,  $\oplus$  represents dilation, and  $\ominus$  represents erosion. In mammography, this method is useful for enhancing lesion boundaries, improving mass segmentation, and refining tumor detection in dense breast tissue. Its strengths include enhancing tumor visibility and improving CADx system segmentation. However, it is highly dependent on the choice of structuring element, and poor parameter tuning can introduce artifacts that affect diagnostic accuracy [79,80].

## 3.2 Statistical methods

Statistical texture descriptors analyze the spatial distribution of pixel intensities to characterize tissue patterns in mammography. These methods capture essential texture properties through statistical measures derived from pixel relationships.

### 3.2.1 First-Order Statistical Methods (Histogram-Based Analysis)

First-order statistics describe texture by analyzing the distribution of pixel intensities without considering spatial relationships. Most important features used in mass description include mean ( $\mu$ ), which represents the average intensity, variance ( $\sigma^2$ ) for intensity dispersion, skewness ( $S$ ) for asymmetry, and kurtosis ( $K$ ) for sharpness of intensity peaks. These are calculated using:

$$\mu = \frac{1}{N} \sum_{i=1}^N I_i \quad (3.15)$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (I_i - \mu)^2 \quad (3.16)$$

$$S = \frac{1}{N} \sum_{i=1}^N \left( \frac{I_i - \mu}{\sigma} \right)^3 \quad (3.17)$$

$$K = \frac{1}{N} \sum_{i=1}^N \left( \frac{I_i - \mu}{\sigma} \right)^4 \quad (3.18)$$

In mass description, these features help differentiate dense and fatty breast tissues, where malignant tumors often exhibit higher variance and skewness due to heterogeneous structures. Strengths include fast computation and ease of implementation, but limitations arise from their inability to capture spatial dependencies, making them insufficient for distinguishing textures with similar intensity distributions [81].

### 3.2.2 Second-Order Statistical Methods

The Gray-Level Co-occurrence Matrix (GLCM) is a powerful second-order statistical method that quantifies spatial relationships between pixel pairs at specific distances and orientations in mammography images. By creating matrices that count how frequently pixel value pairs occur in defined spatial relationships (typically specified by distance and angle), GLCM captures critical textural properties like homogeneity, contrast, correlation, and energy

through Haralick features (e.g., energy, entropy, homogeneity) [81]. In breast imaging, GLCM excels at characterizing tissue patterns by detecting subtle variations in pixel relationships that correlate with malignancy status, smooth margins (typically benign) produce high homogeneity and energy values with low contrast measures, while spiculated or irregular margins (often malignant) generate higher contrast, entropy, and lower homogeneity values. Despite sensitivity to preprocessing choices and the need for careful parameter selection, GLCM remains one of the most effective texture descriptors for discriminating between benign and malignant masses in clinical computer-aided diagnosis systems, with reported accuracy rates exceeding 85% when combined with appropriate classification methods [82,83].

### 3.3 Model-Based Texture Descriptors

Model-based texture descriptors characterize textures using mathematical models that describe patterns and structures within an image. These methods assume that textures follow specific generative models and estimate parameters to describe texture properties effectively. The most common model-based approaches include Markov Random Fields (MRF) and Autoregressive Models (AR).

#### 3.3.1 Markov Random Field (MRF) Models

MRF describe texture by modeling the spatial dependencies between pixel intensities using probabilistic graphical models. MRF assumes that the intensity of a pixel depends on its neighboring pixels, making it useful for capturing local texture structures. The probability of an image configuration is given by:

$$P(I) = \frac{1}{Z} e^{-\sum_c V_c(I)} \quad (3.19)$$

where  $P(I)$  is the probability of the intensity configuration,  $V_c(I)$  represents the potential function modeling pixel interactions within a clique  $c$ , and  $Z$  is a normalization constant. In mammography, MRF is used for mass detection and tissue segmentation. Strengths of MRF model include its ability to capture complex spatial dependencies, making it effective for detecting architectural distortions. However, its limitations include high computational cost and the need for parameter tuning, which can make implementation challenging [84].

#### 3.3.2 Autoregressive (AR) Models

Descriptors based on AR models describe texture by assuming that the intensity of a pixel is a weighted sum of its neighboring pixel intensities, plus a noise term. The AR model is mathematically represented as:

$$I(i,j) = \sum_m \sum_n a_{m,n} I(i-m, j-n) + \epsilon(i,j) \quad (3.20)$$

where  $a_{m,n}$  are the autoregressive coefficients, and  $\epsilon(i,j)$  is the noise term. In mammography, AR models help characterize mass texture by quantifying regularity vs. randomness, with benign masses typically revealing smooth, predictable textures, while malignant tumors have

irregular, unpredictable patterns due to heterogeneity. AR models have the ability to model local intensity variations efficiently, but they are limited in capturing highly irregular and spiculated mass structures, which are often indicative of malignancy [85].

## 3.4 Transform Models

Transform models analyze texture by converting an image from the spatial domain to a different domain, such as frequency or scale-space, where distinct textural patterns associated with benign and malignant masses can be effectively characterized. These methods help in mass detection, classification, and segmentation by highlighting key structural differences in mammographic images. In the literature most used transform models in mass description include the Curvelet Transform (CT), Wavelet Transform (WT), and Gabor Filters.

### 3.4.1 Wavelet Transform (WT)

The Wavelet Transform (WT) is a multi-scale analysis technique that decomposes an image into different frequency components while preserving spatial information. It enables detailed texture analysis of masses by capturing both coarse and fine details. The Discrete Wavelet Transform (DWT) is mathematically defined as:

$$W(j, k) = \sum_i I(i)\psi_{j,k}(i) \quad (3.21)$$

Where :

- $I(i)$  represent the image intensity values.
- $\psi_{j,k}(i)$  are wavelet basis functions at scale  $j$  and position  $k$ .

The Wavelet Transform (WT) is widely used in mass texture description in mammography due to its ability to analyze multi-scale texture variations. Benign masses typically show low-frequency dominance, appearing smooth and homogenous, whereas malignant masses generate high-energy wavelet coefficients across multiple scales, reflecting their heterogeneous and irregular structure [87]. WT is particularly effective in spiculation detection, an important indicator of malignancy. Its main advantages include the ability to capture texture at multiple scales, preserve both spatial and frequency information, and increase lesion segmentation, making it useful for identifying tumor margins. However, WT is computationally expensive, requiring significant processing power, and its effectiveness depends on the choice of wavelet basis functions (e.g., Haar, Daubechies). Additionally, highly noisy mammograms may reduce its performance, necessitating preprocessing steps to improve accuracy.

### 3.4.2 Curvelet Transform (CT)

The Curvelet Transform is a multi-scale directional transform method designed to capture curvilinear features more effectively than the Wavelet Transform. It is particularly useful for analyzing complex textures and structures such as spiculations and mass boundaries in mammographic images. The Continuous Curvelet Transform is mathematically represented by:

$$C(a, b, \theta) = \int_{x \in \mathbb{R}^2} f(x) \psi_{a,b,\theta}(x) dx \quad (3.22)$$

Where :

- $f(x)$  represents the image intensity values.
- $\psi_{a,b,\theta}(x)$  are curvelet basis functions at scale  $a$ , location  $b$ , and orientation  $\theta$ .

The Curvelet Transform has good performance in capturing curvilinear features and spiculations, preserving directional information for improved mass boundary detection and enabling robust multi-scale, multi-orientation analysis that aids in tumor classification. However, its high computational expense, noise sensitivity, and complex implementation can limit its real-time application in mammography. Overall, the Curvelet Transform is a powerful tool for describing mass texture, particularly in detecting irregular, spiculated tumor boundaries, making it a valuable for breast cancer diagnosis in CADx systems [86].

### 3.4.3 Gabor filters

Gabor Filters are transform-based methods defined by a Gaussian-modulated sinusoidal function that simultaneously localizes frequency and spatial information, making them highly effective for capturing texture and shape details. Mathematically, they are expressed as:

$$G(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos(2\pi f x' + \phi) \quad (3.23)$$

where  $x'$  and  $y'$  are the rotated coordinates,  $f$  represents the frequency of the sinusoid,  $\sigma$  determines the width of the Gaussian envelope,  $\gamma$  is the spatial aspect ratio, and  $\phi$  is the phase offset [88]. In the context of mammographic mass description with a focus on shape, these filters effectively highlight structural features such as contours, spiculated edges, and subtle boundary variations by analyzing the image at multiple scales and orientations. This detailed extraction of directional information aids in distinguishing benign masses, which typically have smooth, regular shapes, from malignant tumors, which often present irregular, spiculated outlines. However, the application of Gabor filters requires careful parameter tuning and is computationally intensive, with performance potentially impacted by image noise and contrast variability.

## 4. Evaluation and characteristics of descriptors

Effective evaluation of descriptors in mammography is critical for ensuring the accuracy and consistency of breast cancer screening, as even minor errors in interpreting imaging data can lead to delayed diagnoses or unnecessary interventions. High-quality descriptors must capture subtle patterns in mammograms, such as microcalcifications or irregular masses, while remaining resilient to variations in breast density, imaging artifacts, and equipment differences

or image rotation and many other transformations [89]. Their design should prioritize computational efficiency to support real-time clinical workflows and seamless integration with diagnostic tools, enabling radiologists to make informed decisions swiftly. Additionally, descriptors must generalize across diverse patient populations and imaging protocols to maintain diagnostic consistency. Finally, robust evaluation frameworks and well-designed descriptors enhance early detection rates, reduce false outcomes, and build clinician trust in AI-assisted or automated systems. In this section we will explore most important properties of good descriptors and evaluation of methods studied above.

## **4.1 Characteristics of good descriptors**

Evaluating image descriptors requires a systematic review of their inherent properties to ensure they reliably and accurately represent object features across diverse imaging conditions. Below is a detailed breakdown of the key characteristics of a good descriptor, with each point expanded for clarity:

### **4.1.1 Scale Invariance**

A scale-invariant descriptor consistently represents features regardless of the object's size or image resolution. This is crucial in medical imaging, where lesions or masses can appear at various scales due to differences in patient anatomy or image acquisition settings. A descriptor that maintains its performance across scales allows for more robust comparison and classification, reducing errors caused by zoom or resolution changes [90].

### **4.1.2 Rotation Invariance**

Rotation invariance ensures that the descriptor accurately characterizes features even if the object's orientation varies. In practice, this means that a lesion or mass will have a similar descriptor regardless of whether the image is taken from a different angle or the tissue is rotated. This property is especially important in mammography, where the orientation of masses can differ between images, enabling consistent analysis without the need for pre-alignment [90].

### **4.1.3 Uniqueness (Distinctiveness)**

A unique or distinctive descriptor provides a clear and unambiguous representation of an object, allowing it to be easily distinguished from other objects or textures. This property is critical for applications like tumor classification, where subtle differences in texture and shape must be detected to differentiate benign from malignant lesions. The more unique the descriptor, the lower the chance of misclassification [89,90].

### **4.1.4 Robustness to Noise**

Robust descriptors remain effective even in the presence of image noise or artifacts, which are common in clinical imaging. Noise robustness is vital for ensuring that the descriptor captures true underlying features rather than random variations, leading to more reliable

diagnostics [89,90]. Techniques such as smoothing or normalization can be incorporated into descriptor design to enhance this property.

#### **4.1.5 Illumination Invariance**

Illumination invariance refers to a descriptor's ability to deliver consistent performance despite variations in lighting conditions or exposure settings. In mammography, differences in tissue density and imaging protocols can lead to significant variations in brightness and contrast. A descriptor that is insensitive to these changes will ensure that the captured features are representative of the tissue properties rather than the imaging conditions [89,90].

#### **4.1.6 Computational Efficiency**

Computational efficiency is essential, particularly for real-time or high-throughput applications like computer-aided diagnosis (CADx) systems. A good descriptor should be computationally lightweight, allowing for fast processing without sacrificing accuracy. Efficiency also facilitates integration into clinical workflows, where timely results can be critical for patient management [89,90].

#### **4.1.7 Discriminative Power**

The discriminative power of a descriptor reflects its ability to capture subtle differences between classes, such as benign versus malignant masses. This is often quantified using metrics like classification accuracy, precision, recall, and the F1-score. A highly discriminative descriptor can effectively highlight variations in texture, shape, and intensity that are critical for accurate diagnosis, thereby improving the performance of automated detection systems [89,90].

#### **4.1.8 Comprehensively Descriptive**

Beyond capturing texture, a comprehensive descriptor should encapsulate additional features such as shape, edges, and spatial relationships. In the context of mammography, this means that the descriptor should provide information about the smoothness or irregularity of mass boundaries, the presence of spiculations, and the overall spatial arrangement of tissue structures. Such a holistic representation facilitates a more complete understanding of the pathology, aiding in more accurate detection and classification [90].

### **4.2 Evaluation**

Below in Table 3.1 is a comparative table summarizing various descriptors including basic shape measures, transform-based, statistical, and model-based descriptors focusing on their features, invariance properties, strengths, and limitations for mass description (texture and shape) in mammography.

<b>Descriptor</b>	<b>Features Captured</b>	<b>Invariance</b>	<b>Strengths</b>	<b>Limitations</b>
<b>Area, Perimeter, Rectangularity, Centroid, Circularity, Compactness</b>	Basic geometric and shape properties (size, boundary, symmetry, and compactness)	Can be made scale- and rotation-invariant with normalization ; sensitive to segmentation accuracy	Simple, intuitive, computationally efficient; directly interpretable	Highly dependent on segmentation quality; limited in capturing texture or internal heterogeneity
<b>Fourier Descriptor</b>	Global shape boundary features represented as frequency components	Can be normalized for translation, rotation, and scale invariance	Robust to minor boundary perturbations; effective global shape representation	Sensitive to noise; less effective in capturing localized details
<b>Moments-Based Descriptors</b> (e.g., Hu, Zernike moments)	Overall shape characteristics computed from image moments	Often designed to be invariant to scale, rotation, and translation	Compact representation; robust global descriptors with proven invariance properties	Can be sensitive to noise and segmentation errors; may require higher-order moments for complex shapes
<b>Fractal Analysis</b>	Self-similarity and complexity of shapes/textures (irregularity and heterogeneity)	Intrinsically scale invariant	Excellent for characterizing irregular, heterogeneous textures typical of malignant masses	Computationally intensive; parameter selection can be challenging
<b>Skeletonization-Based Descriptors</b>	Medial axis and topological structure (structural “skeleton” of the shape)	May be normalized, but highly sensitive to the quality of	Emphasizes structural topology and connectivity; useful for identifying	Sensitive to noise and segmentation errors; skeleton extraction can be unstable

		the extracted skeleton	branching/spiculated patterns	
<b>Histogram-Based Analysis</b>	Global intensity distribution (first-order statistical features)	Not inherently invariant; normalization can provide partial invariance	Simple, fast, and computationally efficient; good for overall texture assessment	Ignores spatial relationships; limited in discriminative power for complex textures
<b>GLCM (Gray-Level Co-occurrence Matrix)</b>	Second-order statistics capturing spatial relationships between pixel intensities	Not directly invariant; rotational invariance achieved by averaging over angles	Highly discriminative for texture analysis; effective in capturing heterogeneity	Computationally intensive; sensitive to parameter (distance, angle) selection
<b>Markov Random Field (MRF) Models</b>	Local spatial dependencies modeled probabilistically	Can be tailored for invariance through model design, but often model-specific	Captures spatial interactions and context; useful for segmentation and pattern recognition	Computationally expensive; requires careful parameter tuning
<b>Autoregressive (AR) Models</b>	Local pixel intensity dependencies and regularities	Can be normalized for partial invariance but not inherently robust to large variations	Simple and efficient for regular textures; useful for differentiating smooth vs. irregular patterns	Less effective for highly irregular textures; sensitive to noise
<b>Wavelet Transform (WT)</b>	Multi-scale frequency and spatial information; captures both	Can be made scale invariant; some degree of rotation invariance	Provides multi-resolution analysis; effective for mass classification and segmentation	Computationally demanding; performance dependent on choice of wavelet basis

	coarse and fine details	with proper design		
<b>Curvelet Transform (CT)</b>	Curvilinear and directional features; ideal for capturing spiculated, elongated structures	Similar invariance to wavelets with enhanced directional sensitivity; normalization possible	Superior for capturing curvilinear edges and irregular boundaries; excels in detecting spiculated tumor margins	Highly computationally expensive; complex implementation and sensitive to noise
<b>Gabor Filters</b>	Multi-scale, multi-orientation texture and shape details; emphasizes edge and contour features	Invariance depends on parameter tuning; can be made rotation invariant through filter bank design	Capture directional patterns and fine details; robust in highlighting spiculated edges	High computational cost; sensitive to noise and requires careful parameter selection

**Table 3.1:** Evaluation of different methods for mass description.

This table provides an overview of how each descriptor performs in terms of the essential properties required for effective mass description. The choice of descriptor or combination thereof will depend on the specific clinical application, the imaging conditions, and the desired balance between computational efficiency and discriminative power.

## 5. Proposed model for breast mass description (PATAR descriptor)

This section aims to introduce a novel descriptor to equip radiologists with a robust descriptor that significantly enhances the diagnostic process within a CADx system. Effective diagnosis relies on a descriptor that can extract critical features; however, the unpredictable morphology of spiculated masses and their high similarity to surrounding breast tissue pose substantial challenges. Many existing descriptors fall short in capturing all spiculated regions, leading to misclassifications [112]. To address this, we propose a two-step approach shape descriptor called PATAR (Polygon Approximation Triangle-Area Representation).

First, we perform a geometric transformation by approximating the mass contour with a polygon, which accurately isolates and estimates the concave and convex regions key characteristics that differentiate benign from malignant masses. Next, we employ the Triangle-Area Representation (TAR) to quantify the spiculation size: by traversing the polygon's vertices

in a clockwise direction, we form triangles from every set of three consecutive points, using their areas as a measure of spiculated extent. Our PATAR descriptor, combining polygon approximation with TAR calculation, offers a robust solution for detecting and sizing spiculations. Figure 3.2 illustrates the PATAR workflow, with detailed explanations provided in the subsequent sections.



**Figure 3.2:** PATAR descriptor process, starting with polygon approximation applied on mass ROI's then TAR signature to extract features generation.

## 5.1 Polygon approximation

Polygonal approximation is a technique that simplifies complex contours by representing them with a polygon composed of a minimal set of vertices. The primary objective is to reduce the level of detail while retaining the essential geometric features of the original shape, such as critical concave and convex regions. This reduction is achieved by selecting a subset of key points along the curve, in that way preserving important structural characteristics while enhancing computational efficiency, data compression, and visualization [113]. Moreover, because the resulting polygon is inherently insensitive to translation, scale, and rotation, it offers robust invariance properties, making it highly effective for shape characterization and classification across various applications. In our work, polygon approximation is particularly important for preserving the degree of spiculation in masses. By reducing the mass contour to a polygonal form, we effectively capture and preserve the relevant features, both concave and convex that are indicative of malignancy.

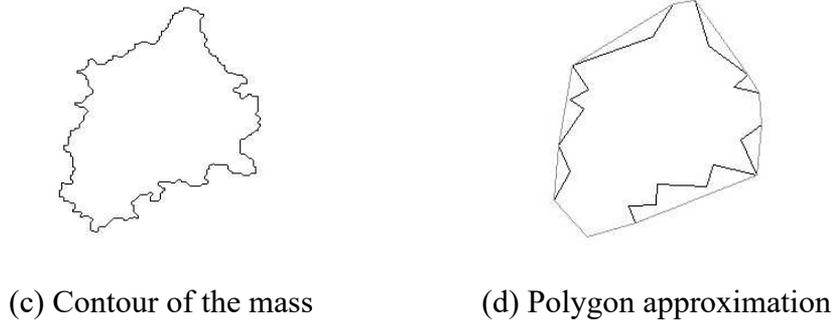
Figure 3.3 illustrates how polygon approximation is applied to a digital mammographic mass, demonstrating its capacity to simplify the contour while keeping the critical spiculated characteristics intact.



(a) Original mammographic mass



(b) CBIS-DDSM ROI



**Figure 3.3:** Example of Polygon approximation: a) the original mammographic mass, b) CBIS-DDSM ROI, c) contour and shape of the mass and d) polygon approximation of the mass.

The Ramer-Douglas-Peucker (RDP) algorithm [113] is employed to achieve polygonal approximation of breast mass contours. Consider a discrete curve  $C_d = \{p_1, p_2, \dots, p_n\}$ , where each point  $p_i = (x_i, y_i)$  represents a coordinate along the mass boundary in a clockwise sequence. These contours are initially extracted from mammographic images through edge-detection methods. The RDP algorithm begins by selecting the first and last points of the curve, designated as the anchor and floating points, respectively. It then computes the perpendicular distance from each intermediate point to the straight line connecting these two endpoints. If the maximum distance exceeds a user-defined threshold  $\epsilon$ , this point is preserved as it indicates a significant deviation from a straight line, and the algorithm recursively processes the sub-curves on either side of this point. The key formula in the RDP algorithm computes the perpendicular distance  $d$  from a point  $p = (x, y)$  to the line defined by endpoints  $p_l = (x_l, y_l)$  and  $p_n = (x_n, y_n)$ :

$$d(p, \overline{p_l p_n}) = \frac{|(y_n - y_l)x - (x_1 - x_n)y + x_n y_1 - y_n x_1|}{\sqrt{(y_n - y_l)^2 + (x_1 - x_n)^2}} \quad (3.24)$$

The parameter  $\epsilon$  is critical: serves as the threshold value against which the computed perpendicular distance is compared. In the RDP algorithm, if the distance  $d(p, \overline{p_l p_n})$  exceeds  $\epsilon$ , the point is retained; otherwise, it is discarded, a lower  $\epsilon$  value results in a polygon that closely follows the original contour, preserving finer details such as subtle spiculations, while a higher  $\epsilon$  value simplifies the curve more aggressively, potentially omitting essential diagnostic features. Figure 3.4 illustrates how varying  $\epsilon$  values affect the contour representation of the same mass. In Algorithm 1, this threshold is referred to as tolerance. The efficiency of the RDP algorithm in breast mass analysis lies in its ability to reduce the computational burden by simplifying complex contours while retaining the critical geometric characteristics needed for accurate shape characterization and classification in CADx systems.

---

```
procedure DouglasPeucker(PointList[1...n], tolerance: real)
```

```
  dmax := 0
```

---

```
index := 0

for i := 2 to n - 1 do

    d := PerpendicularDistance(PointList[i], Line(PointList[1], PointList[n]))

    if d > dmax then

        index := i

        dmax := d

    end for

if dmax > tolerance then

    recResults1 := DouglasPeucker(PointList[1...index], tolerance)

    recResults2 := DouglasPeucker(PointList[index...n], tolerance)

    return concatenate(recResults1, recResults2)

else

    return Line(PointList[1], PointList[n])

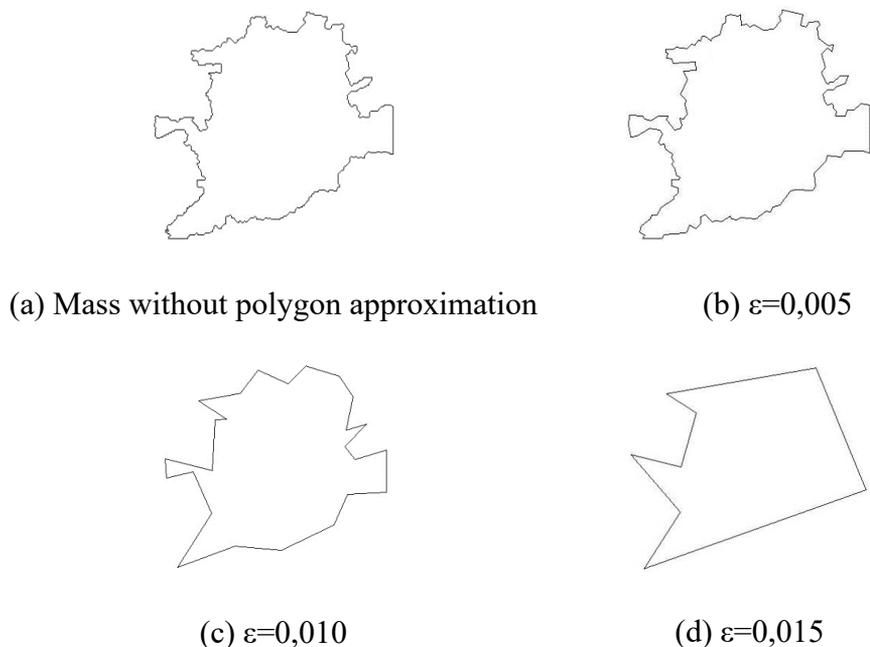
end if

end procedure
```

---

**Algorithm 3.1:** Douglas-Peucker polygon approximation algorithm

The value of the  $\epsilon$  parameter is fundamental for the PATAR descriptor; a false value may omit the concave and convex spaces in mass. The main challenge lies in selecting an appropriate value for the  $\epsilon$  parameter in the polygon approximation process: it must be optimal enough to preserve the detailed morphology of spiculated regions without introducing distortion, while also effectively simplifying the contour of regular masses into a polygonal shape that retains their essential characteristics. In this study, the  $\epsilon$  was tested with many values around the universal value of 0.01. This universal setting has been found to provide an optimal balance, ensuring accurate contour representation while maintaining the morphological integrity of the masses.



**Figure 3.4:** Different  $\epsilon$  values applied to the same mass contour: (a) the original mass contour without approximation; (b) with  $\epsilon = 0.005$ ; (c) with  $\epsilon = 0.010$ ; and (d) with  $\epsilon = 0.015$

## 5.2 Triangle-area Representation (TAR) signature

The Triangle Area Representation (TAR) signature is a geometric signal processing technique that transforms a one-dimensional data sequence into a feature representation by calculating the signed areas of triangles formed by consecutive points. By leveraging computational geometry principles, TAR computes triangle areas using three successive signal points, creating a unique signature that captures local signal variations, curvature, and inflection points. This method offers computational efficiency, scale invariance, and rotation independence, making it particularly useful in domains like pattern recognition, biomedical signal analysis, fault diagnosis, and texture characterization. The transformation provides a compact geometric representation that preserves essential signal characteristics while reducing computational complexity, enabling advanced feature extraction and analysis across various applications.

Concavity and convexity, formed by spiculations in masses, are critical features for differentiating malignant from benign lesions. In our approach, these features are quantified in the second step of the PATAR descriptor, following polygon approximation, using the Triangle-Area Representation (TAR). This method measures local curvature along the mass contour by computing the area of triangles formed by three consecutive contour points. Specifically, for every three pixels  $P_{n-t_s}=(x_{n-t_s},y_{n-t_s})$ ,  $P_n=(x_n,y_n)$ , and  $P_{n+t_s}=(x_{n+t_s},y_{n+t_s})$ ; where  $t_s$  indicates the step (with  $t_s=1$  meaning that  $P_{n+1}$  and  $P_{n-1}$  are immediate neighbors in the clockwise and counterclockwise directions, respectively), the TAR function is defined as:

$$TAR(n, ts) = \frac{1}{2} \begin{vmatrix} i_{n-ts} & j_{n-ts} & 1 \\ i_n & j_n & 1 \\ i_{n+ts} & j_{n+ts} & 1 \end{vmatrix} \quad (3.25)$$

For every three consecutive pixels of the contour of masses  $P_{n-ts}(i_{n-ts}, j_{n-ts})$ ,  $P_n(x_n, y_n)$ , and  $P_{n+ts}(x_{n+ts}, y_{n+ts})$ , where  $n \in [1, N]$  and  $t_s \in [1, N/2 - 1]$ ,  $t_s$  represent the step of the TAR function if  $t_s=1$  means that  $P_{n+1}$  is the neighbor of  $P_n$  in the clockwise direction and  $P_{n-1}$  is the neighbor in counter-clockwise direction. The triangle is formed by the points  $P_{n-1}$ ,  $P_n$ , and  $P_{n+1}$  is given by Alajlan et al. [106]. A pseudo-code of the TAR signature is presented in Algorithm 3.2.

---

```
Procedure TAR(PointList[1...n], step: integer)
```

```
for p := 1 to n do
```

$$TAR[p] := \det[(p_{i-step}, p_{j-step}, 1), (p_i, p_j, 1), (p_{i+1}, p_{j+1}, 1)]$$

```
if TAR(p) < 0 then p is concave point
```

```
if TAR(p) > 0 then p is convex point
```

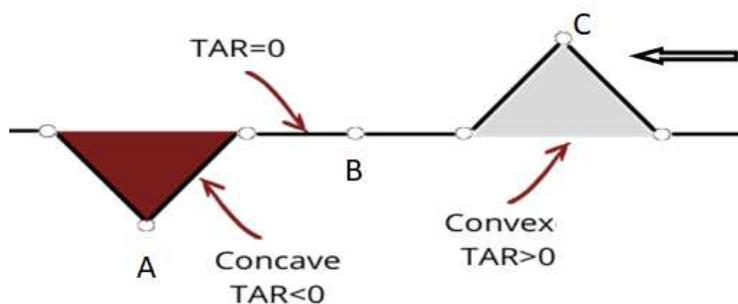
```
if TAR = 0 traight line
```

```
return TAR[]
```

```
end procedure
```

---

**Algorithm 3.2:** TAR signature of contour



**Figure 3.5:** Three possible outcomes for the TAR signature are observed: the contour is convex when TAR is positive, concave when TAR is negative, and forms a straight line when TAR equals zero.

Figure 3.5 demonstrates that the TAR function yields three distinct outcomes negative, zero, and positive when the contour is traversed counter-clockwise. These results reflect the characteristics of the area defined by three consecutive points, indicating whether the region is concave, linear, or convex [114]. Moreover, the TAR signature exhibits remarkable invariance, maintaining its consistency under translation, rotation, and scaling transformations. This combination of efficiency and invariance makes the triangle-area representation an invaluable

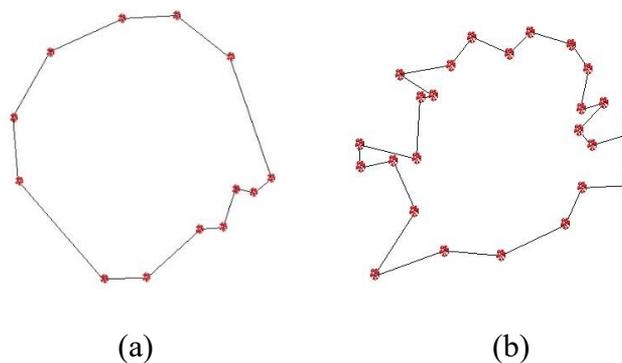
tool in our research, as it accurately renders smooth curves while reducing computational complexity.

### 5.3 Feature extraction

In computer vision and image processing, a descriptor serves to quantitatively and mathematically capture and represent the essential characteristics or features of an image. By simulating human perception, descriptors aim to translate visual features and patterns into measurable data, mirroring the way humans interpret and comprehend visual content. A feature, in this context, refers to quantifiable information extracted from an image to facilitate object identification and classification. The PATAR descriptor specifically targets the quantification of spiculation in masses located in the ROI image, assessing its degree to categorize the mass as either benign or malignant.

The malignant mass shows a unique topology, defined by lines of varying lengths and thicknesses radiating from its margin. As previously mentioned, PATAR is designed to isolate both concave and convex regions by first applying a polygon approximation to the mass contour. All vertices of the resulting polygon are marked to form triangles, whose areas are computed using the Triangle-Area Representation (TAR). The vertices are traversed in a clockwise direction, and for each point  $P_i$ , the area of the triangle formed by  $P_i, P_{i-1}$ , and  $P_{i+1}$  is determined. The extracted features from the PATAR descriptor include the total number of vertices as well as the counts of concave and convex regions.

- **Number of corners:** Highly irregular shapes typically exhibit more corners compared to smoother, round shapes. Oval and circular masses commonly benign tend to have fewer corners than irregular forms. Figures 3.6 and 3.7 illustrate the correlation between the number of corners and the likelihood of a mass being round.

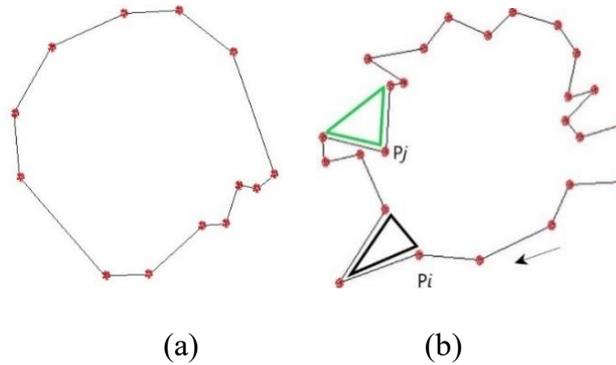


**Figure 3.6:** Number of corners as feature explain clear variance between (a) Benign mass with round shape having 13 corners, and (b) malignant mass with irregular shape having 25 corners.

- **Negative and Positive TAR:** The TAR signature is employed to extract and evaluate the convex and concave regions formed by the polygon's vertices. Specifically, the areas of the triangles generated at each vertex are divided into two groups: those with negative values, summed as  $TARN$ , and those with positive values, summed as  $TARP$ , as indicated in the next equation:

$$TARN = \sum_{i=1}^n TAR_i < 0, \quad TARP = \sum_{i=1}^n TAR_i > 0 \quad (3.26)$$

where  $n$  represents the total number of corners. These aggregated values quantify the degree of spiculation in the mass. High values of both  $TARN$  and  $TARP$  suggest a highly irregular (and likely malignant) mass, while a round mass is characterized by a  $TARN$  value close to zero.



**Figure 3.7:** Illustration of the distinction between a round mass (a) and a spiculated mass (b) based on the values of TARN (green triangles) and TARP (black triangles).

In addition to TARN, TARP, and the number of corners, features such as mass area and contour length are also integrated into the classification process. These extracted features are combined into the PATAR vector: [TARN, TARP, NBP, MA, CL].

## 5.4. Mass classification

The features extracted during the description phase serve as inputs for the classification step, where each mass is assigned to benign or malignant category. Given the high stakes in breast cancer diagnosis, where every decision allows significant consequences and only a biopsy can provide definitive confirmation, we opted for a fuzzy classifier to begin with. Fuzzy classifiers, such as Fuzzy C-means (FCM), assign membership probabilities to each class, offering a probabilistic approach that reflects the inherent uncertainty in medical imaging. In addition, we incorporate Support Vector Machine (SVM) and Random Forest (RF) classifiers to improve the evaluation of the PATAR descriptor further. SVM is chosen for its strong performance in high-dimensional spaces and its ability to define clear decision boundaries, even when the dataset is limited, thereby enhancing classification precision. Meanwhile, RF,

an ensemble learning method, is adept at handling heterogeneous data and mitigating overfitting, which improves the model's generalization capabilities. By utilizing these three classifiers, FCM, SVM, and RF, we ensure a robust and comprehensive evaluation of the PATAR descriptor, leveraging both probabilistic and discriminative approaches tailored to the data's complexity and the critical nature of breast cancer diagnosis.

### 5.4.1 Fuzzy Clustering

Fuzzy C-means (FCM) is a soft clustering algorithm that assigns each data point a degree of membership across several clusters rather than forcing a hard assignment to a single cluster. In other words, each data point is associated with a probability score indicating its likelihood of belonging to each cluster [115]. Although fuzzy classification is not commonly used in CADx systems, it has shown promise in breast cancer diagnosis because it reflects the uncertainty in the classification of masses where only a biopsy can provide a definitive diagnosis, and radiologists often rely on probabilities when evaluating malignancy.

Fuzzy C-means (FCM) is typically an unsupervised clustering algorithm, but it can be adapted for supervised classification by incorporating labeled training data. In this approach, class centroids are first computed from the labeled dataset, which provides a reference for each class. New samples are then assigned membership probabilities based on their distances to these centroids. This supervised adaptation allows FCM to utilize prior knowledge to improve classification accuracy. As a result, each sample receives a probability score indicating its likelihood of belonging to each class. Thus, FCM can effectively function as a fuzzy classifier in a supervised setting.

The membership probability of a mass is determined by the distance between each feature vector (representing an image) and the centroid of the respective class (malignant or benign). Based on the ground truth training dataset,  $N$  masses are divided into two classes, malignant and benign  $J_a$  and  $J_b$ , respectively, and the matrix membership  $U_{ia}, U_{ib} \in \{0,1\} \ i=1, \dots, N$  is also created. For each class, the centroid is calculated as follows [115]:

$$C_{a,b} = \frac{\sum_{i=1}^N U_{i(a,b)}^m x_i}{\sum_{i=1}^N U_{i(a,b)}^m} \quad (3.27)$$

In the classification stage, the distance from each mass's feature vector to the centroids of the benign and malignant classes is computed. These distances are then used to estimate the membership of each mass to both classes. The resulting membership values,  $U_{i(a,b)}$ , indicate the probability of a mass being either malignant or benign.  $U_{i(a,b)}$  is calculated as follows [115]:

$$U_{i(a,b)} = \frac{1}{\sum_{k=1}^c \left( \frac{\|x_i - c_{a,b}\|}{\|x_i - c_{k,a,b}\|} \right)^{\frac{2}{m-1}}} \quad (3.28)$$

### 5.4.2 Support Vector Machine

A Support Vector Machine (SVM) is a supervised machine learning algorithm designed for classification and regression tasks. Its primary goal is to identify the optimal hyperplane a decision boundary in a high-dimensional space that maximizes the margin between distinct classes. The margin is defined as the distance between the hyperplane and the closest data points (called support vectors) from each class. These support vectors are critical, as they determine the hyperplane's position and orientation. For non-linearly separable data, SVM employs the kernel trick, which maps input features into a higher-dimensional space (e.g., using linear, polynomial, or radial basis function kernels) to enable effective separation. By focusing on support vectors and margin maximization, SVM ensures robust generalization and minimizes overfitting.

Key parameters of SVM include the regularization parameter ( $C$ ), which controls the trade-off between maximizing the margin and minimizing classification error, and the kernel function, which defines how the data is transformed into a higher-dimensional space. Common kernels include the linear, polynomial, and radial basis function (RBF) kernels, each with its own set of parameters (e.g., gamma for the RBF kernel, degree for the polynomial kernel). Additionally, parameters such as tolerance for stopping criteria and epsilon in regression tasks influence the optimization process. Together, these parameters allow SVMs to be finely tuned to balance model complexity, training accuracy, and computational efficiency, making them a versatile tool in various application domains, including medical diagnostics and image classification.

### 5.4.3 Random Forest Classifier

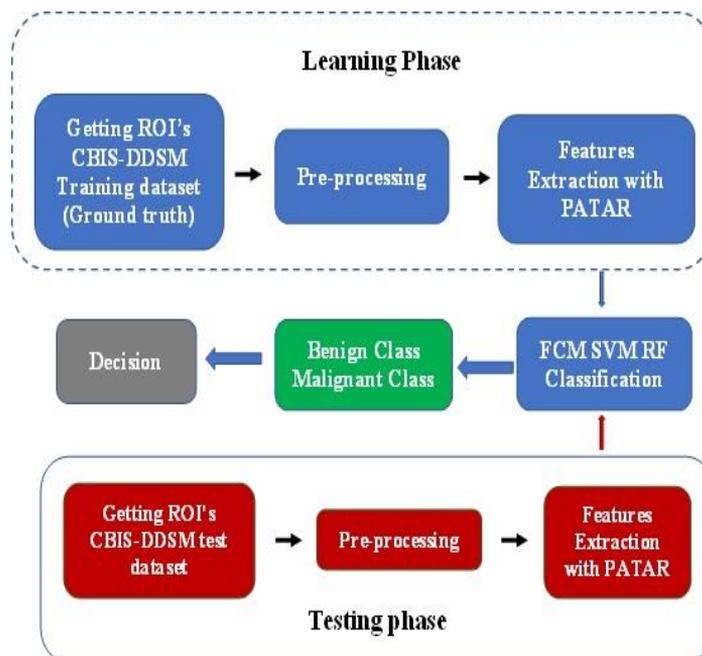
Random Forest is an ensemble learning method that builds multiple decision trees during training and merges their outputs to generate a robust classification. By aggregating the predictions of many individual trees each constructed using a random subset of the data and features this approach reduces overfitting and improves generalization. It is well-suited for handling complex, high-dimensional data and can be used for both classification and regression tasks, making it a versatile choice in various domains, including medical diagnostics.

Key parameters of the Random Forest classifier include the number of trees ( $n\_estimators$ ), which determines how many decision trees are built; the maximum depth ( $max\_depth$ ) of each tree, which limits the complexity of the model; and the minimum number of samples required for a split, among other settings. In our approach, the classifier's hyperparameters were tuned via grid search to achieve optimal performance on our dataset, ensuring a balance between variance reduction and maintaining high classification accuracy for breast mass analysis.

## 5.5 Experiments and results

The adopted strategy in experiments is separated into two phases: learning and testing. In the training part, features are extracted from the ROIs of the CBIS-DDSM dataset. ROIs are a portion of the images containing the abnormality; these are delimited and annotated by mammographs and radiologists in the CBIS-DDSM dataset. CBIS-DDSM is an improved and

standardized version of DDSM designed for evaluating CAD systems [108]. Using the ground truth of the CBIS-DDSM dataset, we evaluate the PATAR descriptor. Figure 3.8 shows the outline of our CADx system based on PATAR.



**Figure 3.8:** The proposed approach builds a feature model using training data, applies SVM, FCM, and RF classifiers to classify malignant vs. benign masses, and evaluates PATAR's diagnostic performance during testing.

Our approach for automatic diagnosis was put to the test through a series of experiments. These experiments evaluated its performance by measuring accuracy, sensitivity, specificity, and the F1-score. Python software (version 3.7) with OpenCV library was used on a PC with Intel i5 (2.00 Ghz) with 8 GB of RAM and a Windows 10 operating system. CBIS-DDSM was used to assess the model's performance [116]. The dataset is separated into two subsets, one for the training containing 1318 ROIs (637 Malignant and 681 benign masses) and another for the test and validation step composed of 378 ROIs (147 Malignant and 231 benign masses).

### 5.5.1 CBIS-DDSM dataset

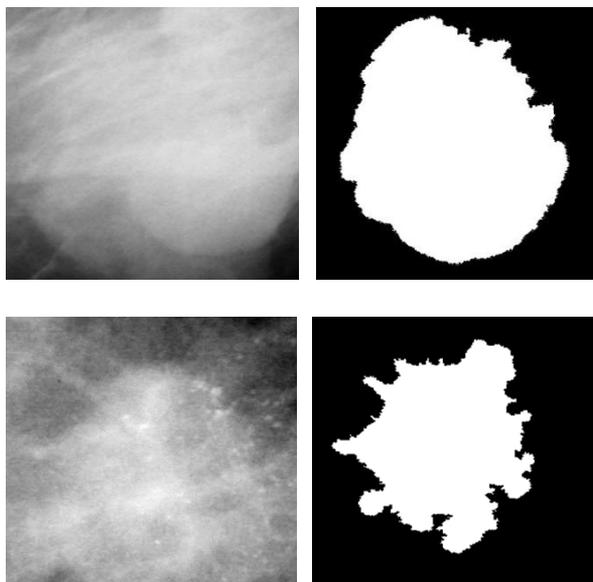
The CBIS-DDSM (Curated Breast Imaging Subset of the Digital Database for Screening Mammography) dataset is a widely used, publicly available resource in breast imaging and computer-aided diagnosis. It is a curated version of the original DDSM dataset, offering high-resolution full-field digital mammography images in DICOM format along with comprehensive annotations that include lesion boundaries and region-of-interest (ROI) markings. These ROI annotations, typically delineated by expert radiologists, are essential for accurately localizing potential abnormalities such as masses and calcifications. The dataset includes cases with both benign and malignant lesions, captured in standard views like mediolateral oblique (MLO) and craniocaudal (CC) [116].

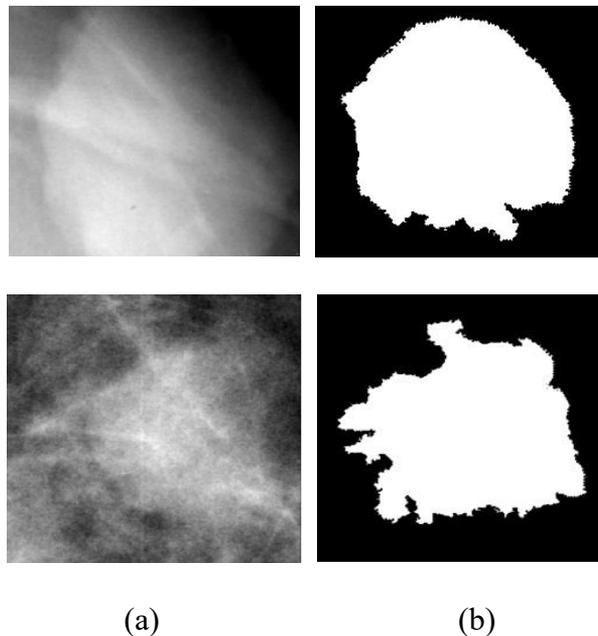
Researchers value the CBIS-DDSM dataset for its detailed ROI annotations and standardized segmentations, which facilitate the precise extraction of key features necessary for the development and benchmarking of advanced machine learning algorithms and CADx systems. The use of DICOM format ensures that the images maintain critical metadata and high image quality, which is crucial for both diagnostic analysis and reproducibility of research. Despite challenges related to image variability and noise, the comprehensive nature of the dataset including its DICOM images and ROI information makes it an invaluable tool for advancing breast cancer research and enhancing the reliability of diagnostic models. Table 3.2 shows the number of cases, abnormalities, and images in each set, training, and testing. Masses are located and approved by an experiment radiologist [165].

	Benign cases			Malignant cases			
	Cases	Abnormalities	Images	Cases	Abnormalities	Images	Total Images
Training	355	387	681	336	361	637	1318
Testing	117	135	231	83	87	147	378

**Table 3.2:** Number of Cases, Abnormalities and Images in the Training and Test Sets, each case can have one or more abnormalities and more images.

Computer-aided diagnosis systems (CADx) need only analyze regions of interest (ROIs), not full mammogram images. ROIs feature abnormalities within the cropped sections of the image, which outline the bounding rectangle of the abnormality relative to its ROI see Figure 3.9. Our descriptor performed calculations directly on masses segmented and delineated the mass from the enveloping tissue. Ground truth provided by CBIS-DDSM is founded on the BI-RADS category.





**Figure 3.9:** Four cropped images of mass from CBIS-DDSM dataset, a) mammographic images without segmentation, b) shows segmented mass (Mask image).

## 5.5.2 Experimentation process

Below is a detailed report outlining the machine learning pipeline that integrates preprocessing, feature extraction using the PATAR descriptor, and subsequent classification employing FCM, SVM, and Random Forest (RF).

### 5.5.2.1 Preprocessing

The initial stage of the pipeline involves preprocessing the digital mammography images. This step is crucial for enhancing image quality and preparing the data for robust feature extraction. Preprocessing typically includes:

- **Conversion and Thresholding:** The grayscale images are converted into binary representations by applying a threshold. This operation segregates the mass from the background based on intensity values, where pixels above the threshold are set to 255 and those below to 0.
- **Morphological Operations:** To refine the binary images, morphological erosion and dilation are applied. Erosion helps remove small artifacts and noise, while dilation restores the original shape of the masses, smoothing their contours.
- **Contour Extraction:** The refined binary images are then used to extract the contours of the masses. The contours, retrieved serve as the basis for subsequent polygon approximation.

### 5.5.2.2 Features extraction (PATAR Descriptor)

Following preprocessing, the PATAR descriptor is employed to capture the essential shape and spiculation characteristics of the mass. This descriptor involves two major steps:

- **Polygon Approximation:** Using the Ramer-Douglas-Peucker (RDP) algorithm, the complex mass contour is approximated by a polygon. This approximation reduces the number of contour points while preserving critical geometrical features such as concave and convex regions. The polygonal representation not only facilitates efficient computation but also maintains invariance to translation, scaling, and rotation. Optimal  $\varepsilon = 0.01$  is used.
- **Triangle-Area Representation (TAR):** Once the mass contour is approximated, the TAR technique is applied. The resulting TAR values are categorized into negative (concave), positive (convex), and zero (flat segments), providing a signature that quantifies the degree of spiculation. Additionally, features such as the number of corners, mass area, and contour length are extracted and compiled into a feature vector: [TARN, TARP, NBP, MA, CL].

### 5.5.2.3 Classification

In the proposed experimentation pipeline all data passed by preprocessing and features extraction steps, but in the classification, data is separated into learning data and test data. In the learning phase, the feature vector is used to train multiple classifiers Fuzzy C-Means (FCM), Support Vector Machine (SVM) with an RBF kernel (with C set to 1.5, determined through grid search), and Random Forest (best parameters are also determined through grid search). The training process involves using a labeled dataset where each mass is classified as benign or malignant, enabling the models to learn the discriminative patterns in the feature space. Once the classifiers are trained, the testing phase begins, where new, unseen mammographic images are preprocessed and their feature vectors are extracted using the PATAR descriptor. These vectors are then fed into the trained classifiers to predict the class membership of each mass, providing probability scores that reflect the likelihood of malignancy or benignity. This dual-stage approach comprising both learning and testing ensures robust performance and reliable classification in breast cancer diagnosis.

### 5.5.3 Results and discussion

To evaluate the efficacy of our proposed approach, we used a total of 1,545 mammograms. Out of these, 1,322 images (685 benign and 637 malignant) were allocated for training, while 243 images (181 benign and 62 malignant) were reserved for testing. The diagnostic performance of our method is evaluated based on computing time, sensitivity, specificity, and accuracy. Accuracy indicates the proportion of test cases that the classifier correctly identifies, sensitivity (SN) reflects the true positive rate, and specificity (SP) denotes the true negative rate. These performance metrics are formally defined as follows:

$$SN = \frac{TP}{TP+FN} \quad (3.29)$$

$$SP = \frac{TN}{TN+FP} \quad (3.30)$$

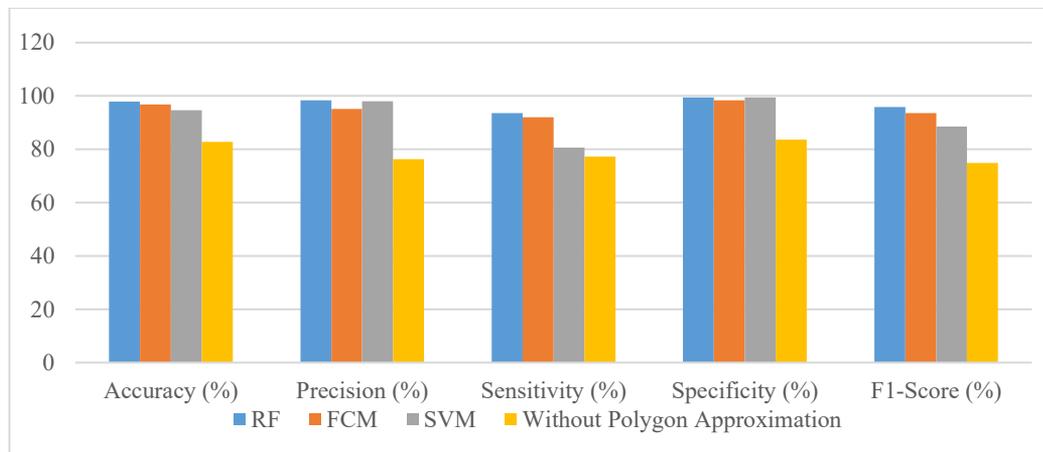
$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (3.31)$$

where, TP is the true positive, FP is the false positive, TN is the true negative, and FN is the false negative.

In medical image context accuracy alone can be misleading, especially in imbalanced datasets common in medical diagnostics, because it may mask the classifier's inability to correctly identify the minority class. For instance, if benign cases dominate, a classifier might achieve high accuracy by simply predicting most cases as benign, while failing to detect malignant cases. This is why sensitivity and specificity are critical: sensitivity (true positive rate) measures how effectively the model identifies actual malignant cases, ensuring that few cancers are missed, and specificity (true negative rate) assesses how well benign cases are correctly classified, reducing false positives. Together, these metrics provide a more nuanced and clinically relevant evaluation of diagnostic performance than accuracy alone.

The evaluation metrics obtained after experimentations demonstrate that integrating polygon approximation with the TAR signature significantly enhances the automatic classification of mammograms for breast cancer diagnosis. Polygon approximation plays a crucial role in our descriptor by simplifying the complex shapes and contours of breast masses especially spiculated ones into a reduced set of vertices without sacrificing key characteristics. This simplification allows the Triangle-Area Representation (TAR) to more effectively quantify spiculations. Without polygon approximation, TAR computations would be performed on all minor contour variations, leading to an inaccurate description and poor estimation of the spiculations.

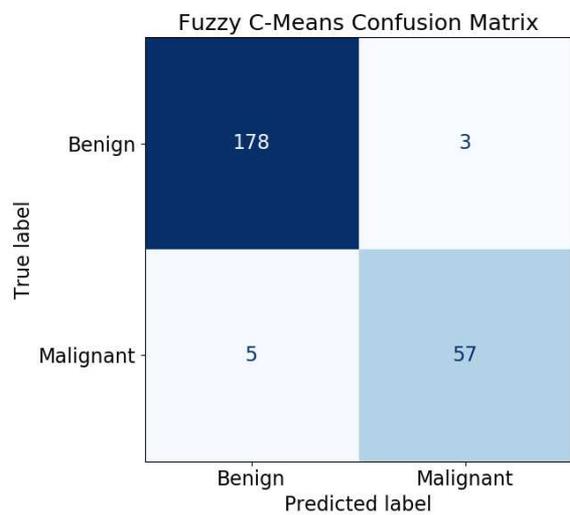
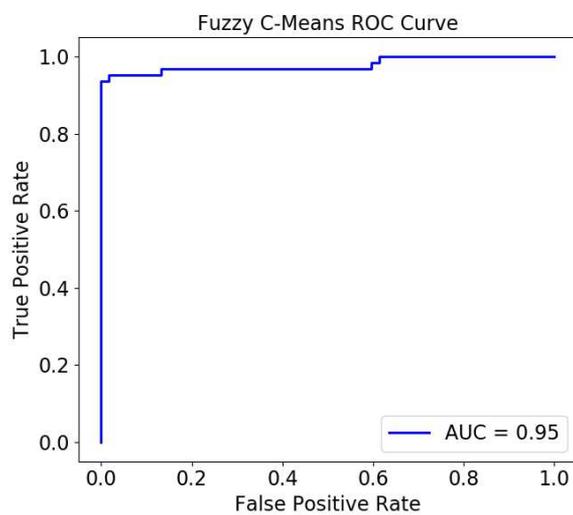
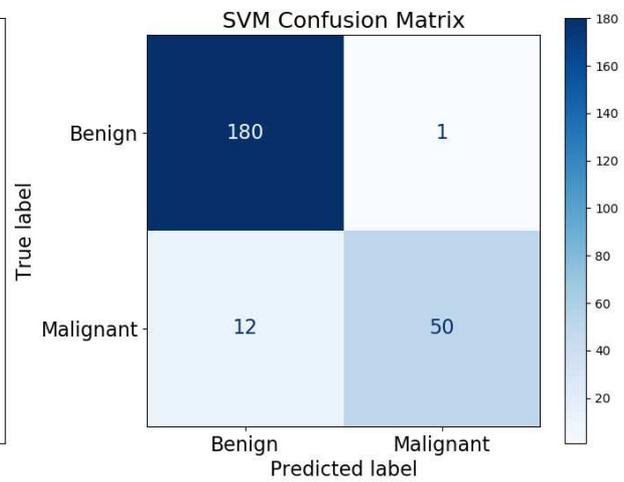
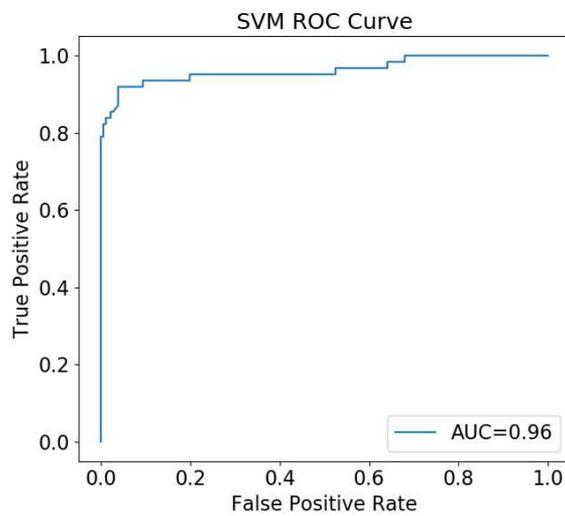
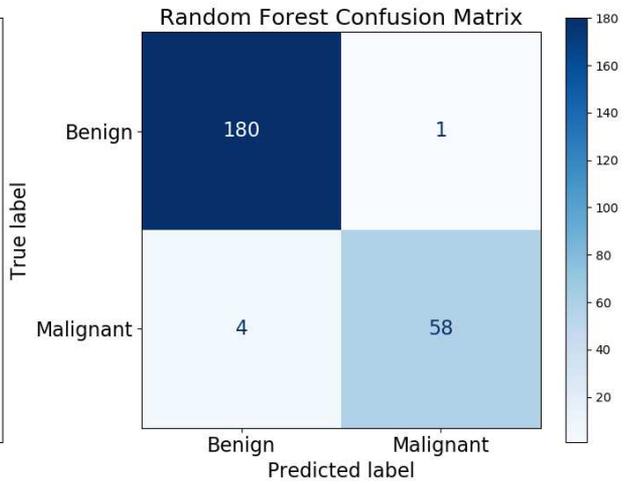
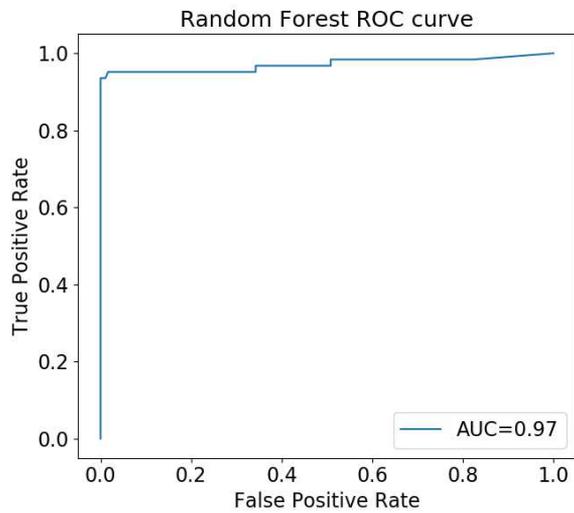
In Table 3.3, the results of the comparison between our approach and classification without polygon approximation using TAR signature. The accuracy obtained with polygon approximation and Fuzzy C-means classifier is 96.76%, 97.94% with Random Forest, and 94.65% with Support vector machines. Without polygon approximation, the classification accuracy decreases significantly to 82.80% [176,177]. The amelioration gained in terms of accuracy, precision, sensitivity, and specificity with polygon approximation confirms with exactitude our hypotheses that spiculation in masses will be well raised, represented, extracted, and evaluated with our contribution using PATAR descriptor. Compared results are presented in Figure 3.10.



**Figure 3.10:** Comparing the best results of PATAR, with and without polygon approximation.

Figure 3.11 shows the ROC plots that illustrate the classification performance of the three classifiers, RF, FCM, and SVM, using the DDSM dataset. The area under the ROC curve (AUC) serves as an aggregate measure of performance across different classification thresholds, with values ranging from 0 to 1, where 1 represents perfect classification. The AUC for the SVM classifier was 95.48, while FCM achieved an AUC of 96.23. Remarkably, the Random Forest classifier outperformed the others, obtaining an AUC of 97.13 and an accuracy of 97.94.

In addition to ROC curve, the confusion matrix illustrated in Figure 3.11, provides valuable insight into the classification performance of the models beyond standard metrics such as accuracy or AUC. In our results, PATAR correctly classified a high number of benign (true negatives), 180 cases with RF and SVM and 178 cases with FCM, and malignant (true positives) cases, with only a few misclassifications. The low number of false negatives (malignant tumors incorrectly predicted as benign) is particularly significant in the context of breast cancer, as such errors could delay critical treatment. Similarly, the minimal false positive rate suggests the model avoids over-diagnosing benign cases as malignant, which could otherwise lead to unnecessary stress or interventions. These observations are further supported by the high sensitivity and specificity values, indicating the model's balanced ability to detect both classes accurately. Overall, the confusion matrix confirms that the Random Forest classifier performs reliably in distinguishing between benign and malignant tumors, making it a promising tool for aiding clinical decision-making.



**Figure 3.11:** ROC curve and confusion matrix of RF, SVM and FCM using CBIS-DDSM. Random Forest classifier ensures the best results in terms of accuracy and AUC.

	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	Execution Time
With Polygon Approximation						
RF	97,94	98,31	93,55	99,45	95,87	26' 35''
FCM	96,76	95,08	92,06	98,37	93,55	24' 17''
SVM	94,65	98,04	80,65	99,45	88,50	25' 51''
Without Polygon Approximation	82,80	76,29	77,32	83,57	74,81	5' 23''

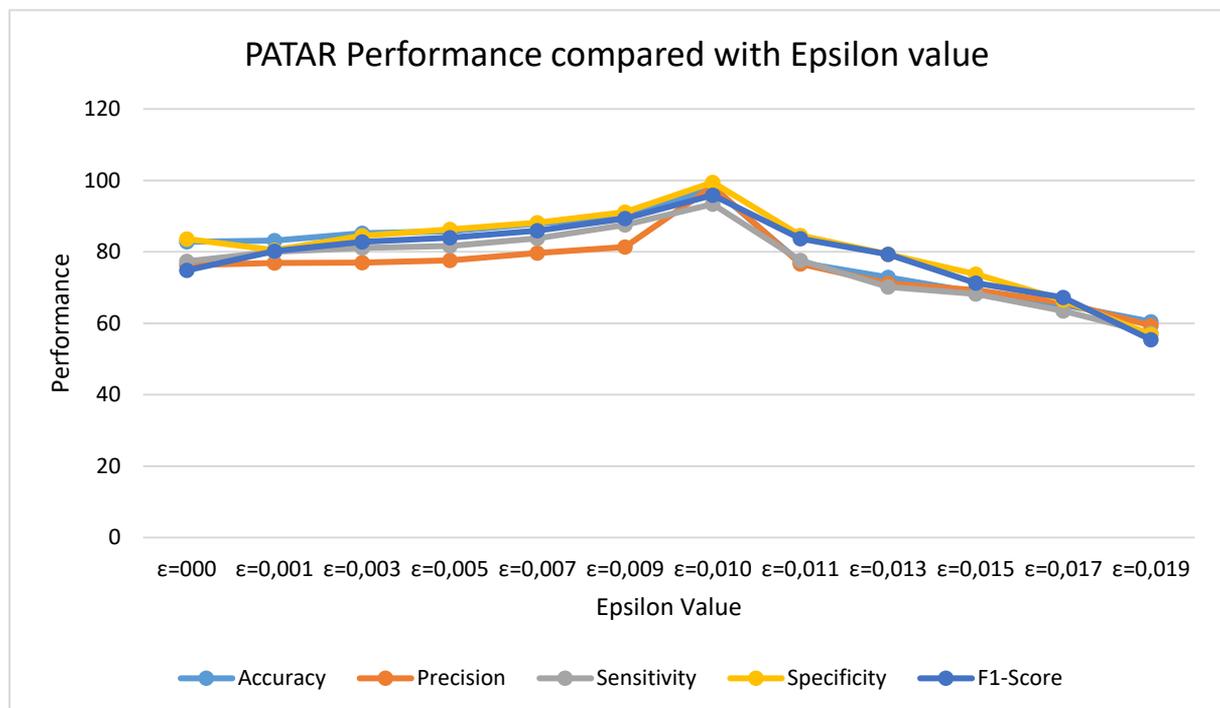
**Table 3.3:** Comparison of results obtained with and without polygon approximation.

The effect of the parameter  $\epsilon$  on the classification performance was thoroughly investigated in Table 3.3, and the results demonstrate a clear sensitivity of the model to this parameter. The parameter  $\epsilon$  in this tests series represents the simplification threshold in the Ramer–Douglas–Peucker (RDP) algorithm, which is applied for polygonal approximation of mass boundaries. Its impact on classification performance is highly significant. As  $\epsilon$  increases from 0.000 to 0.010, the performance metrics show a steady improvement, suggesting that moderate simplification helps reduce noise in contour and highlights more relevant structural features. At  $\epsilon = 0.010$ , the model achieves its highest performance, with an accuracy of 97.94%, precision of 98.31%, sensitivity of 93.35%, specificity of 99.45%, and F1-score of 95.87%. These values confirm that optimal boundary simplification through polygon approximation enhances the model's ability to differentiate between benign and malignant cases. However, further increasing  $\epsilon$  beyond this optimal point leads to a steep decline in all metrics as shown in Figure 3.12. This suggests that excessive simplification results in the loss of critical contour details, which negatively affects the model's discriminatory capacity. Therefore,  $\epsilon$  must be carefully selected to maintain the balance between eliminating noise and preserving meaningful geometric information. This emphasizes the crucial role of the RDP parameter in preprocessing steps for breast mass classification.

	Accuracy	Precision	Sensitivity	Specificity	F1-Score
$\epsilon=000$	82,8	76,29	77,32	83,57	74,81
$\epsilon=0,001$	83,14	76,85	79,995	80,42	80,21
$\epsilon=0,003$	85,21	76,99	81,1	84,52	82,81
$\epsilon=0,005$	85,75	77,56	81,66	86,24	83,95

$\epsilon=0,007$	87,9	79,68	83,79	88,12	85,96
$\epsilon=0,009$	89,8	81,34	87,57	91,09	89,33
$\epsilon=0,010$	<b>97,94</b>	<b>98,31</b>	<b>93,35</b>	<b>99,45</b>	<b>95,87</b>
$\epsilon=0,011$	76,95	76,66	77,59	84,59	83,67
$\epsilon=0,013$	72,83	71,17	70,17	79,33	79,28
$\epsilon=0,015$	68,69	69,29	68,16	73,78	71,27
$\epsilon=0,017$	65,18	65,80	63,47	66,48	67,23
$\epsilon=0,019$	60,46	59,35	57,22	56,83	55,38

**Table 3.4:** PATAR performance using Random Forest classifier with different values of Epsilon.



**Figure 3.12:** Evolution of PATAR performance cross different values of Epsilon, all metrics decreased when epsilon exceeds 0.01 the universal value.

### 5.5.3.1 Comparison with state-of-the-art

In this section, PATAR descriptor is compared with some of most powerful feature-based methods in the stat-of-the-art. Table 3.4 presents detailed description of methods. The highest accuracy was achieved by the PATAR with RF (Random Forest) classifier, reaching 97.94% on the CBIS-DDSM dataset. Other top-performing methods include PATAR-FCM (96.76%) and Xie et al. (96.02%), demonstrating the effectiveness of feature extraction and classification strategies. Methods such as de Brito Silva et al [109]. (93.70%) and Pezeshki et al. (93.22%)

also performed well, while Beheshti et al [101]. (87.81%) had the lowest accuracy, suggesting that fractal-based lesion discrimination may be less effective than shape and texture-based methods.

Study	Method Description
<b>De Brito Silva et al. [109]</b>	Proposed a descriptor utilizing geometric and topological features derived from two spatial feature maps: the distance map and the surface map. These maps capture mass geometry and topology while preserving spatial information. Shape descriptors based on distance histograms help compute distances, areas, and angles.
<b>Pezeshki et al. [108]</b>	Developed a method to extract spiculated pixels from tumors with uniform intensity. Pixel similarity is evaluated by measuring dissimilarities between a central pixel and its neighbors across multiple symmetric orientations. Small differences indicate spiculations, which are then extracted by summing dissimilarity values across all directions.
<b>Vijayarajeswari et al. [107]</b>	Utilized the Hough transform to identify and isolate specific shapes in mammographic images. This method is effective for detecting various shapes and straight lines, demonstrating resilience against noise, gaps, and occlusions. An accumulator function is computed for each edge point, facilitating precise feature extraction.
<b>Souza et al. [102]</b>	Introduced a method leveraging shape descriptors (D1dist, D2dist, and D3dist) along with convolution-based analysis. By selecting random surface pixels and segmenting the object along the z-axis, statistical shape features such as mean and standard deviation are extracted, enhancing mass classification.
<b>Beheshti et al. [101]</b>	Applied a fractal-based approach to distinguish lesions from background tissues. This technique minimizes mean square error while extracting asymmetric lesion information. The method focuses on defining new fractal features for malignancy assessment based on the region of interest (ROI).

**Table 3.5:** Comparison of Methods for Mammographic Mass Characterization.

In terms of sensitivity and specificity, different methods exhibit varying levels of performance. Beheshti et al. (97.37% sensitivity, 79.55% specificity) focuses on detecting malignant tumors effectively but at the cost of higher false positives. Conversely, PATAR – RF (93.55% sensitivity, 99.45% specificity) and PATAR – FCM (92.06% sensitivity, 98.37% specificity) provide a balanced approach with high specificity, reducing false positives while

maintaining reliable detection of malignant cases. Pezeshki et al. (92.06% sensitivity, 94.54% specificity) also achieves a strong balance, making it a competitive method.

The choice of classifier plays a critical role in the effectiveness of mammographic mass classification. Support Vector Machine (SVM) is the most commonly used classifier, applied in 7 out of 10 methods, proving its reliability for mass detection. However, Random Forest (RF) in PATAR – RF achieved the highest accuracy, indicating that ensemble methods may provide better performance in feature learning. LDA (Linear Discriminant Analysis) and Fuzzy logic classifiers, used in Rabidas et al. and Goudarzi et al., respectively, showed moderate performance, suggesting that traditional statistical approaches may be less effective than modern machine learning models.

METHOD	Nb. images	Database	Accuracy (%)	Sensitivity (%)	Specificity (%)	Classifier	Year
PATAR – FCM [168]	1545	CBIS-DDSM	96.76	92.06	98.37	FCM	2024
PATAR – SVM [168]	1545	CBIS-DDSM	94.65	80.65	99.45	SVM	
PATAR – RF [168]	1545	CBIS-DDSM	97.94	93.55	99.45	RF	
Arora et al. [110]	-	CBIS-DDSM	88	-	-	SVM	2020
de Brito Silva et al. [109]	794	DDSM	93.70	96.29	91.05	High Boost	2020
Kaur et al [118]	322	MIAS	94.8	93	90	KNN	2019
Pezeshki et al. [108]	200	DDSM	93.22	92.06	94.54	SVM	2019
Vijayarajeswari et al. [107]	322	MIAS	94	-	-	SVM	2019
Rabidas et al [117]	558	MIAS/DDSM	94.57	-	-	LDA	2018
Goudarzi et al. [105]	-	mini-MIAS	89.37	88.23	84.23	Fuzzy	2018
Souza et al. [102]	620	DDSM	92.15	91.40	92.90	SVM	2017
Xie et al. [100]	300	DDSM	96.02	94.88	97.16	SVM	2016
Beheshti et al. [101]	168	DDSM	87.81	97.37	79.55	SVM	2016

**Table 3.6:** Performance comparison between proposed approach and some previous works.

The dataset used significantly impacts the model's performance, with larger datasets generally yielding better results. CBIS-DDSM and DDSM are the most frequently used databases, contributing to high accuracy scores due to their large number of labeled mammographic images. Studies using smaller datasets, such as MIAS or mini-MIAS, tend to

have slightly lower accuracy, as seen in Goudarzi et al. (89.37%), indicating that a limited number of training samples may restrict the model's ability to generalize.

Overall, PATAR descriptor outperforms other methods in accuracy, specificity, and sensitivity, demonstrating the effectiveness of advanced classification techniques. Deep feature-based classifiers, such as RF and FCM, provide a better result compared to traditional classifiers like SVM or KNN. Larger datasets, particularly CBIS-DDSM, contribute to improved model performance, whereas smaller datasets like MIAS show moderate results. These observations suggest that future research should focus on hybrid models combining different feature extraction methods and deep learning techniques to further enhance classification accuracy.

Table 3.6 presents a comparative results analysis between our approach and several existing methods for mammographic classification, clearly highlighting the performance advantages of our method.

## 6. Conclusion

In this chapter, we explored various methods for describing breast masses in mammographic images, highlighting their significance in the accurate diagnosis of breast cancer. Also, we introduced a novel descriptor that increases the characterization of mass shapes, aiming to improve classification performance. This proposed descriptor addresses limitations observed in existing methods by providing a more detailed and discriminative analysis of mass features. The proposed approach starts with a polygon approximation (geometric transformation) on breast mass, then, Triangle-area representation is calculated to evaluate the degree of spiculation of the mass providing to the classifiers robust features and getting high evaluation of the descriptor. The subsequent sections presented experimental results demonstrating the effectiveness of this new descriptor in distinguishing between benign and malignant breast masses. To improve PATAR descriptor, exploring more datasets is recommended. Also, PATAR descriptor can be used for many abnormalities like architectural distortions or microcalcifications.

# Chapter 4: Automatic diagnosis of microcalcifications using Deep Learning

## 1. Introduction

Deep learning has emerged as a transformative force in medical imaging, enabling automated and highly accurate analysis of complex medical images [119]. By leveraging multi-layered neural networks, deep learning models can extract intricate patterns from imaging data, often surpassing human performance in specific diagnostic tasks. These techniques are particularly valuable in radiology, pathology, and other imaging-intensive medical fields, where they assist in detecting abnormalities, quantifying disease progression, and improving diagnostic consistency. Unlike traditional machine learning approaches that rely on manual engineering features, deep learning models learn directly from pixel data, allowing them to capture subtle and complex features that may be imperceptible to human observers or conventional algorithms [120].

One of the most significant applications of deep learning in medical imaging is disease detection and diagnosis. For example, convolutional neural networks (CNNs) have demonstrated remarkable success in identifying tumors in MRI, mammography, and Computed Tomography scans, classifying diabetic retinopathy from retinal images, and detecting pneumonia in chest X-rays. These models can process vast amounts of imaging data rapidly, serving as valuable decision-support tools for clinicians. Additionally, deep learning excels in image classification tasks, such as delineating organ boundaries for radiation therapy planning or isolating lesions for precise measurement such as breast masses. Many models like ResNet, DenseNet or EfficientNet, with encoder-decoder structure, skip connections or bottleneck block, has become a gold standard for medical image segmentation and classification due to its ability to preserve spatial details while learning hierarchical features [121].

Ensemble deep learning refers to the technique of combining multiple deep learning models to improve overall performance, robustness, and generalization compared to using a single model. Stacking or stacked generalization is a powerful ensemble learning technique that integrates diverse models by training a meta-learner, also known as a meta-model to combine their outputs. Our first contribution in this chapter consists in designing of optimal meta-learner using multilayer perceptron to integrate predictions from three different models. The proposed meta-learner plays a crucial role in refining predictions by learning how to weight and integrate the decisions from the base models.

Another ensemble learning technique is Boosting achieved with algorithms like Adaboost and Gradient Boost. The main idea behind boosting is to learn different models sequentially, and adjust weight of samples dataset after each model to make the next model focusing more on misclassified samples. While boosting methods like AdaBoost and Gradient Boosting are commonly employed in ensemble learning, the automatic diagnosis of microcalcifications in medical imaging demands new, specialized algorithms to address

the problem of high false positive and false negative rates. Our second contribution in this chapter includes a new algorithm named Cost-Sensitive Boosting with Error Weighted Adjustments (CSB-EWA). The key innovation in CSB-EWA is the penalties  $\lambda_{FP}$  and  $\lambda_{FN}$  used in weight adjustments. CSB-EWA define weights using false positive and false negative rates, to push models increasing sensitivity and specificity.

This chapter is structured into three main sections. Section 2, provides a comprehensive overview of deep learning, covering fundamental concepts, various architectures, and the challenges encountered in their application, particularly in medical image analysis and microcalcifications. Section 3, presents ensemble strategies in deep learning with most techniques used. Section 4, presents our two proposed approaches for microcalcification classification based on two strategies of ensemble learning, stacking and boosting. Experiments and results will be presented using CBIS-DDSM.

## 2. Deep Learning

Deep learning (DL) has fundamentally transformed medical imaging, driving advancements in diagnostic accuracy, workflow efficiency, and personalized treatment. By leveraging large datasets and powerful neural networks, DL automates tasks that were traditionally manual, time-consuming, and prone to human error. Below, we explore, concepts, architecture of deep learning and its key applications, persistent challenges in automatic diagnosis of breast cancer abnormalities, and future innovations.

### 2.1. A brief history

The origins of deep learning date back to eight decades when a computational model inspired by the neural networks of the human brain was developed. Since then, artificial intelligence has undergone continuous advancement, interrupted only by two major pauses in its progress. These interruptions coincided with the well-known periods of stagnation in artificial intelligence research, commonly referred to as the AI winters [122].

#### 2.1.1 The Birth of AI (1940s – 1950s)

The foundations of artificial intelligence were laid in 1943 when Warren McCulloch and Walter Pitts introduced the first mathematical model of artificial neurons, known as threshold logic. This model attempted to simulate the way human brain neurons process information. In 1950, Alan Turing proposed the Turing Test, a benchmark for evaluating a machine’s ability to exhibit intelligent behavior. A year later, Marvin Minsky and Dean Edmonds built SNARC, the first artificial neural network-based computer. The formal recognition of AI as a distinct field came in 1956 when John McCarthy coined the term “Artificial Intelligence” at the historic Dartmouth Conference, marking the beginning of AI research [122].

#### 2.1.2 Early AI and the First AI Winter (1957 – 1970s)

In 1957, Frank Rosenblatt developed the Perceptron, the first neural network capable of learning, sparking excitement about AI’s potential. This optimism led to the development of early natural language processing systems like ELIZA in 1966, a chatbot

designed to mimic human conversations. However, despite initial progress, AI research faced limitations due to the lack of computational power and unrealistic expectations. By 1973, funding and interest in AI decreased, leading to the First AI Winter, a period of reduced progress and investment that lasted until the 1980s.

### **2.1.2 Neural Networks & The Second AI Winter (1980s – 1990s)**

AI research experienced a revival in the 1980s with the rise of expert systems, rule-based programs used in industries for decision-making. In 1986, Geoffrey Hinton, David Rumelhart, and Ronald Williams introduced backpropagation, a breakthrough algorithm that significantly improved the training of neural networks. Nevertheless, in spite of these advancements, AI systems remained expensive and computationally demanding. As a result, the 1990s saw another decline in AI research, leading to the Second AI Winter, as industries and governments lost confidence in AI's practicality.

### **2.1.3 Rise of Deep Learning (2000s – 2010s)**

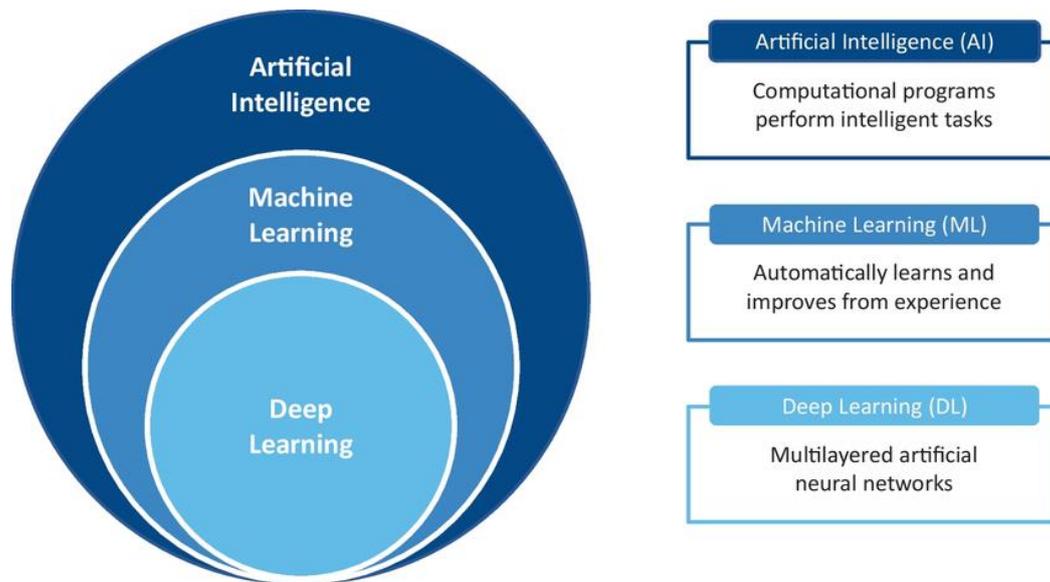
AI gained renewed momentum in the 2000s with the emergence of deep learning, a term popularized by Geoffrey Hinton in 2006. This technique, which used multi-layered neural networks, allowed AI to achieve unprecedented levels of accuracy in tasks such as image and speech recognition. In 2011, IBM's Watson defeated human champions in the quiz show *Jeopardy!*, demonstrating AI's growing capabilities. The breakthrough moment came in 2012 when AlexNet, a deep convolutional neural network (CNN), won the ImageNet challenge, revolutionizing computer vision. AI further advanced in 2014 when Google DeepMind's AlphaGo defeated professional Go players, showcasing the power of reinforcement learning.

### **2.1.4 Present: AI Revolution (2020s)**

The 2020s have been marked by rapid AI advancements, particularly in natural language processing (NLP) and generative AI. In 2023, OpenAI released GPT-4, significantly improving AI's ability to generate human-like text and assist in various applications. AI is now deeply integrated into multiple industries, including healthcare, autonomous systems, finance, and creative fields. Ongoing research focuses on enhancing AI's efficiency, interpretability, and ethical considerations, ensuring responsible and beneficial development for the future.

## **2.2 Definitions**

Today, artificial intelligence (AI), machine learning (ML), and deep learning (DL) are widely used terms that are often mistakenly interchanged to describe intelligent systems or software capable of mimicking human-like decision-making. However, as shown in Figure 4.1 these concepts, while closely related, have distinct meanings and evolutionary paths [123].



**Figure 4.1:** Relation between Artificial Intelligence, Machine Learning and Deep Learning.

### 2.2.1 Artificial Intelligence (AI)

AI is a broad and interdisciplinary field focused on enabling machines or systems to perceive their environment, reason through information, make decisions, and adapt their behavior in a way that mimics human intelligence. This includes the ability to process data, recognize patterns, learn from experience, and improve performance over time without explicit programming. The ultimate goal of AI is to create intelligent systems capable of autonomous decision-making, problem-solving, and human-like interaction across different domains.

### 2.2.2 Machine Learning (ML)

Machine Learning (ML) is a specialized branch and subset of artificial intelligence that empowers machines to extract knowledge from data and learn from it independently. Instead of being explicitly programmed for every task, ML systems use algorithms to detect patterns, make predictions, and continuously improve their performance based on the data they encounter.

At its fundamental, ML includes feeding large amounts of data into algorithms, which then analyze and identify underlying structures or trends. This process allows the system to not only understand the data at a superficial level but also to adapt its behavior over time. As a result, these systems become more accurate and efficient in performing tasks such as image recognition, speech processing, and decision-making [123].

### 2.2.3 Le Deep Learning (DL)

DL is a specialized branch of artificial intelligence derived from machine learning, wherein a system is capable of autonomous learning rather than only executing pre-programmed codes. It leverages artificial neural networks, which are computational models inspired by the structure and function of the human brain [119]. These networks consist of

numerous layers often dozens or even hundreds of interconnected neurons, with each successive layer processing and interpreting the outputs from the previous one. For instance, a deep learning system might first acquire the ability to recognize individual characters before advancing to the recognition of words in a text, or it may initially detect the presence of a face in an image prior to identifying the individual depicted.

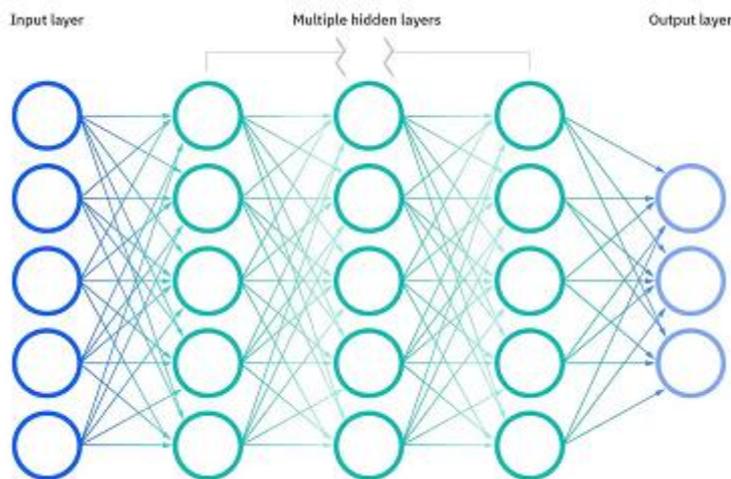
## 2.3 Neuronal networks concepts

Deep learning models use artificial neural networks with an input layer (receives raw data), hidden layers (process data via weighted computations and activation functions), and an output layer (produces predictions). Each neuron applies weights, biases, and nonlinear activations (e.g., ReLU) to transform data hierarchically. The network learns by adjusting weights via backpropagation to minimize prediction errors. This enables automatic feature extraction from complex data like images or text. The result is a system that can classify, predict, or generate outputs without explicit programming [124].

### 2.3.1 Architecture Layers

The architecture of deep learning models is built around interconnected layers that transform input data into meaningful outputs. As shown in Figure 4.2 there are 3 types of layers, input, hidden layers, and output layer.

- **Input Layer:** serves as the entry point for data into the neural network. It is responsible for receiving the raw input or features that the model will process and analyze. Each neuron within this layer corresponds to a specific feature of the input data, such as pixel values in an image, words in a text, or numerical attributes in a dataset. The number of neurons in the input layer depends on the complexity and dimensionality of the input data. This layer does not perform any computations or transformations; instead, it simply passes the data forward to the next layer, ensuring that the neural network can begin learning from the provided information [123,124].



**Figure 4.2:** Architecture of Deep neuronal network, composed of input layer, multiple hidden layers and output layer.

- **Hidden Layers:** these layers are responsible for processing and transforming the input data through a series of computations. Each neuron in a hidden layer receives inputs from the neurons in the preceding layer, applies a set of weighted connections along with a bias, and then processes the result using a mathematical function known as an activation function (e.g., ReLU, Sigmoid, or Tanh). This activation function introduces non-linearity, allowing the network to learn complex patterns and relationships in the data. The processed output is then passed to the neurons in the next layer, enabling the network to progressively refine and extract meaningful features. The depth and number of neurons in the hidden layers play a crucial role in the network's ability to recognize intricate patterns and make accurate predictions [125].
- **Output Layer:** the output layer is the final stage of a neural network, responsible for generating the model's predictions or decisions based on the processed data from the hidden layers. The number of neurons in this layer is determined by the specific task the network is designed to perform. For instance, in binary classification problems (e.g., predicting whether an email is spam or not), a single neuron with a *Sigmoid* activation function is typically used to output a probability value between 0 and 1. In multi-class classification (e.g., identifying different objects in an image), the output layer contains multiple neurons, each corresponding to a distinct class, often using the *Softmax* activation function to assign probabilities to different categories. For regression tasks, where the goal is to predict a continuous value, the output layer usually consists of a single neuron with a linear activation function. The choice of activation function and the number of neurons in this layer directly influence the network's ability to produce meaningful and interpretable results [125].

### 2.3.2 Connections

In an artificial neural network, each neuron within a given layer is systematically connected to all neurons in the adjacent layers. These connections serve as the fundamental pathways through which information is transmitted and processed. Each connection is associated with a specific weight, which quantitatively determines the influence of one neuron on another. During the training process, the neural network employs backpropagation, an optimization technique that iteratively adjusts these weights to minimize errors and enhance predictive accuracy [125,126]. This process is typically facilitated by gradient-based optimization algorithms, such as stochastic gradient descent (SGD) or Adam, which refine the model's parameters by reducing the discrepancy between predicted and actual outputs. The structure and adaptability of these connections are crucial in enabling the network to learn complex patterns and generalize effectively to unseen data [127].

### 2.3.3 Activation Functions

Activation functions are mathematical functions and plays important role in neural networks by introducing non-linearity, allowing models to learn complex patterns and

relationships in data. Without activation functions, the entire network would behave as a simple linear model [128]. Four commonly used activation functions are ReLU, Sigmoid, Softmax, and Leaky ReLU, each with distinct mathematical properties and applications. The choice of activation function in medical imaging depends on the specific task. ReLU and Leaky ReLU are ideal for deep convolutional networks used in segmentation and feature extraction, while Sigmoid and Softmax are better suited for classification problems in disease detection and automatic diagnosis.

- **Rectified Linear Unit (ReLU):** this function is defined as  $f(x) = \max(0, x)$ , meaning it outputs zero for negative inputs and remains linear for positive values. This simplicity makes ReLU computationally efficient and helps prevent the vanishing gradient problem, which can slow down learning in deep networks. However, one major drawback is the dying ReLU problem, where some neurons output only zero and stop learning entirely [129].
- **Sigmoid:** this function, expressed as  $f(x) = \frac{1}{1+e^{-x}}$ , maps any real-valued input into the range (0,1), making it useful for probability estimation in binary classification tasks. It is smooth and differentiable, but suffers from the vanishing gradient problem, where very large or very small inputs produce extremely small gradients, leading to slow learning in deep networks [121].
- **Softmax:** for multi-class classification problems, the Softmax function is often used. It is defined as  $f(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$ , where  $x_i$  represents the input to the  $i^{\text{th}}$  neuron, and the denominator is the sum of exponentials over all inputs in a layer. Each output represents a probability distribution across multiple classes. Softmax ensures that all outputs sum to one, making it ideal for assigning class probabilities. However, it can be computationally expensive due to its reliance on exponentials and is sensitive to large input values, which may lead to numerical instability [122].
- **Leaky ReLU:** A variation of ReLU, addresses the issue of inactive neurons by allowing small nonzero gradients for negative inputs. It is defined as  $f(x) = \max(\alpha x, x)$ , where  $\alpha$  is a small constant, typically set to 0.01. Unlike standard ReLU, Leaky ReLU ensures that neurons continue to learn even when their inputs are negative, making it a more stable alternative in deep networks. However, selecting an appropriate value for  $\alpha$  requires careful tuning [129].

Each activation function has specific strengths and weaknesses, making them suitable for different types of neural network architectures. ReLU and its variants are widely used in deep learning due to their efficiency, while Sigmoid and *Softmax* are commonly applied in classification problems. The choice of activation function significantly impacts the learning process and overall performance of a neural network.

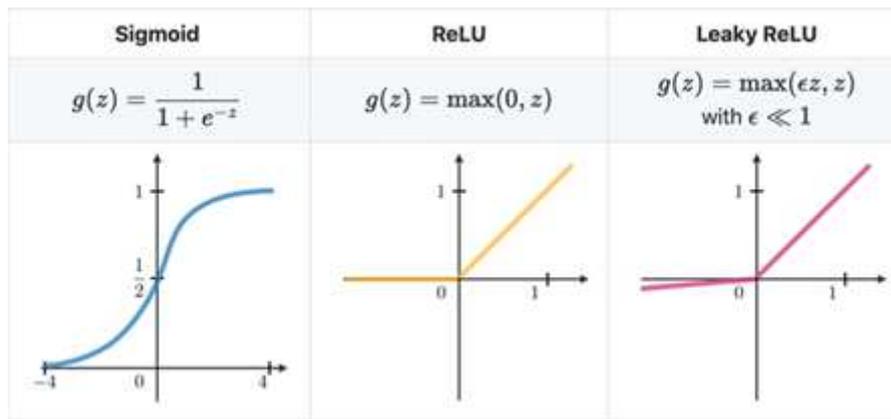


Figure 4.3.:

Figure 4.3: Graphic representation of Sigmoid, ReLU and Leaky ReLU

### 2.3.4 Training

Neural networks acquire knowledge through a process known as training, to make accurate predictions by adjusting its internal parameters. These parameters, known as weights and biases, are fine-tuned so that the network can learn patterns from input data and produce the desired outputs. The training process involves repeatedly exposing the network to data, evaluating its performance, and updating its parameters to minimize errors [124].

The training process consists of several key steps. First, during forward propagation, input data passes through the network, and each layer applies mathematical operations to transform the data via the activation functions mentioned in the sections below. The final output is compared to the true value using a loss function, which measures how far the prediction is from reality. Next, backpropagation calculates how much each weight contributed to the error by computing gradients (derivatives of the loss with respect to each weight). Finally, an optimization algorithm (such as SGD or Adam) updates the weights to reduce the loss in the next iteration [131].

Several important concepts influence how training is conducted. The learning rate determines how much weights are adjusted in each update, too high, and the model may overshoot optimal values; too low, and training becomes slow. Training is typically done in epochs, where each epoch represents one full pass through the dataset. To improve efficiency, data is often split into batches, allowing the model to update weights incrementally rather than all at once. Techniques like dropout and batch normalization help prevent overfitting and stabilize training [132].

Training deep neural networks comes with several challenges. Vanishing gradients occur when gradients become too small to make meaningful updates, often in deep networks with sigmoid or tanh activations, ReLU and residual connections help mitigate this. Overfitting happens when the model memorizes training data instead of generalizing, which can be addressed using regularization techniques like L2 weight decay

or early stopping. Additionally, choosing the right optimizer (e.g., Adam, RMSprop) significantly impacts training speed and final performance [132].

### 2.3.5 Prediction

Once the training process is complete, the neural network is capable of making predictions on previously unseen data by propagating the new inputs through its layers and generating an output from the final layer. This process is called inference and relies on the learned weights and activation functions, which have been optimized during training to capture relevant patterns and relationships within the data [126, 131].

Fundamentally, a neural network functions as a pattern recognition system, learning to map input features to corresponding outputs by iteratively refining its internal parameters. Through exposure to a diverse set of training examples, the network adjusts its weights to minimize errors, thereby enhancing its ability to generalize beyond the training set. This capacity for generalization enables the network to apply learned representations to novel data, making it a powerful tool for predictive modeling in various domains, including image recognition, natural language processing, and time series forecasting [126, 131].

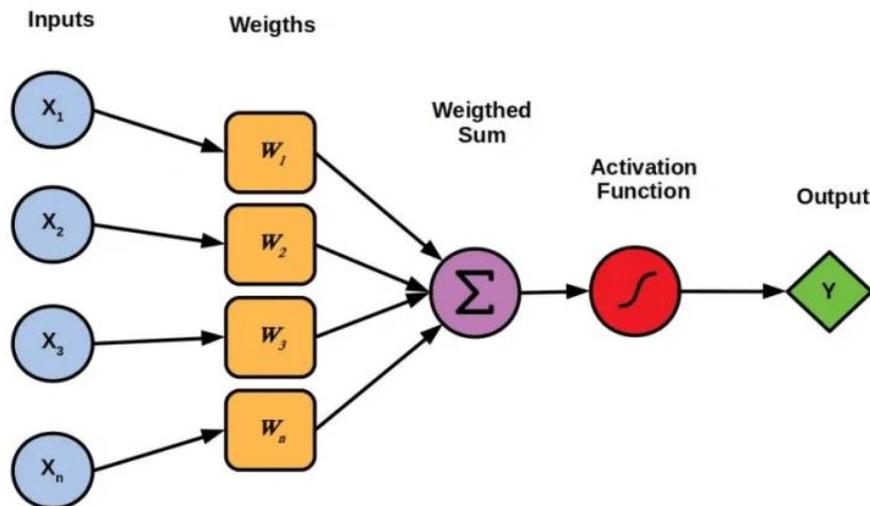
## 2.4 Different Types of Neural Networks in Deep Learning

Neural networks vary in structure and function, each suited for specific tasks. Convolutional Neural Networks (CNNs) excel in spatial feature extraction, making them ideal for tumor detection, organ segmentation, and microcalcification classification. Recurrent Neural Networks (RNNs) and LSTMs handle sequential imaging, while Transformers process long-range dependencies in radiology reports. Autoencoders, GANs, and GNNs aid in anomaly detection, data augmentation, and complex medical relationship modeling, with hybrid models enhancing diagnostic accuracy. In what follows, major types of artificial networks are discussed.

### 2.4.1 Perceptron

The Perceptron is the simplest type of artificial neural network and serves as the foundation for more complex architectures. It is a feedforward model that consists of a single layer of neurons with adjustable weights and biases. The Perceptron operates by computing a weighted sum of input features, applying an activation function (typically a step function or sign function), and producing a binary output (0 or 1) [133].

- **Single-layer Perceptron (SLP):** is a type of neural network that consists of an input layer directly connected to an output neuron, with no hidden layers in between. It works by calculating a weighted sum of the input features, applying an activation function, and producing a binary output. The SLP is capable of solving linearly separable problems, such as AND, and OR logic gates, where a straight line can separate the two classes. However, its simplicity makes it limited in handling more complex, non-linear data, as it cannot effectively model patterns that are not linearly separable, such as XOR logic, which requires more advanced architectures like multi-layer perceptrons (MLPs) [133].



**Figure 4.4:** Architecture of a single layer perceptron (SLP)

- **Multi-layer Perceptron (MLP):** is a type of neural network that consists of multiple hidden layers between the input and output layers, allowing it to model more complex relationships within the data. Unlike the Single-layer Perceptron (SLP), which can only solve linearly separable problems, an MLP uses non-linear activation functions such as ReLU or sigmoid in the hidden layers, enabling it to handle non-linearly separable problems. This makes MLPs capable of learning intricate patterns in data, such as image classification or speech recognition. The network is trained using backpropagation, where the error is propagated backward through the network to update the weights, and gradient descent is employed to minimize the loss function and optimize the model's parameters. MLPs are widely used in various applications, including medical imaging, where they can classify complex data patterns like tumors or organ structures [134].

## 2.4.2 Feedforward Neural Networks (FNNs) Network

Also known as feedforward networks, are a type of shallow neural network where the connections between the nodes do not form any cycles. In these networks, data moves in one direction from input to output, through one or more layers of nodes, each transforming the data based on its weights and activation functions. The learning process in feedforward networks involves adjusting the weights during training to minimize the error between the predicted and actual target values, similar to the process in a perceptron. This adjustment is achieved through algorithms like backpropagation and gradient descent [135].

### Applications of FNNs

- **Facial Recognition:** FNNs are foundational in facial recognition systems, processing large volumes of image data to identify and verify faces even in noisy conditions.
- **Natural Language Processing (NLP):** These networks are widely used for tasks like **speech recognition** and **text classification**, enabling machines to understand and process human language.

- **Computer Vision:** In computer vision, FNNs play a significant role in image classification and object detection, automating the interpretation and analysis of visual data.

### Limitations of FNNs

- **Lack of Feedback Connections:** FNNs do not have feedback connections, meaning they are unsuitable for tasks that require previous outputs to influence future outcomes, such as in sequence prediction problems.
- **Difficulty with Temporal Sequences:** FNNs struggle to model time series data or problems where the sequence of inputs over time is essential.

While FNNs have limitations in handling sequential data or problems requiring temporal dependencies, they have laid the groundwork for more sophisticated models, such as recurrent neural networks (RNNs) and transformers. Despite these constraints, they remain highly effective in solving a wide range of real-world problems, from image classification to natural language understanding.

### 2.4.3 Radial Basis Function (RBF) Neural Network

A Radial Basis Function (RBF) Neural Network is a type of artificial neural network that uses radial basis functions as its activation functions. It is a type of feedforward network typically composed of three layers: an input layer, a hidden layer with RBF units (such as Gaussian functions), and an output layer. The network operates by transforming input data into a higher-dimensional space using radial basis functions, which allow the network to capture non-linear relationships [136].

The Radial Basis Function (RBF) Neural Network consists of three layers: an input layer, a hidden layer, and an output layer. The input layer passes data to the hidden layer, where each neuron calculates the Euclidean distance between the input and a prototype vector (center). The RBF function (typically Gaussian) is then applied to this distance, producing a smooth, localized response that diminishes as the input moves farther from the center. The output layer combines these responses linearly to generate the final prediction. RBF network training involves selecting the centers and widths of the RBF units, often using clustering methods like k-means, and training the weights from the hidden to the output layer with linear regression or other optimization techniques [136].

### Applications of RBF Networks

- **Function Approximation:** RBF networks are often used for function approximation tasks, where they can model complex, non-linear functions.
- **Classification:** RBF networks are effective for classification tasks, especially when the data is separable in a transformed feature space.
- **Time-Series Prediction:** While not as commonly used for sequential data as RNNs or LSTMs, RBF networks have been applied in time-series forecasting, especially when data exhibits clear, localized patterns.

### Limitations of RBF Networks

- **Sensitivity to Parameters:** The performance of RBF networks heavily depends on the choice of parameters, such as the center and width of the RBFs, which can be difficult to tune.
- **Overfitting:** With too many RBF units (centers), the network may overfit the training data, especially in high-dimensional spaces.
- **Scalability:** RBF networks can become computationally expensive when dealing with large datasets, especially in terms of clustering and distance calculations.

#### 2.4.4 Recurrent Neural Network (RNNs)

Recurrent Neural Networks (RNNs) are a class of deep neural networks designed specifically to handle sequential data, excelling in tasks like speech recognition, natural language processing, and time series prediction. Unlike Feedforward Neural Networks (FNNs), which process inputs independently, RNNs are distinguished by their memory mechanism, allowing them to retain information from previous inputs to influence the current output. This is achieved through feedback connections in the hidden layer, where the output from the hidden layer is fed back into itself, creating an internal state that helps the network learn from past sequences. This unique structure enables RNNs to understand and process sequential data, such as text or speech, where the order of the data points is essential [137]. The hidden state stores information from previous time steps and is updated at each sequence step using the formula:

$$h_t = f(W_h h_{t-1} + W_x x_t + b)$$

where:

- $h_t$  is the hidden state at time step  $t$ ,
- $h_{t-1}$  is the hidden state from the previous time step,
- $x_t$  is the input at the current time step,
- $W_h$  and  $W_x$  are weight matrices,
- $b$  is the bias term, and
- $f$  is an activation function, typically tanh or ReLU.

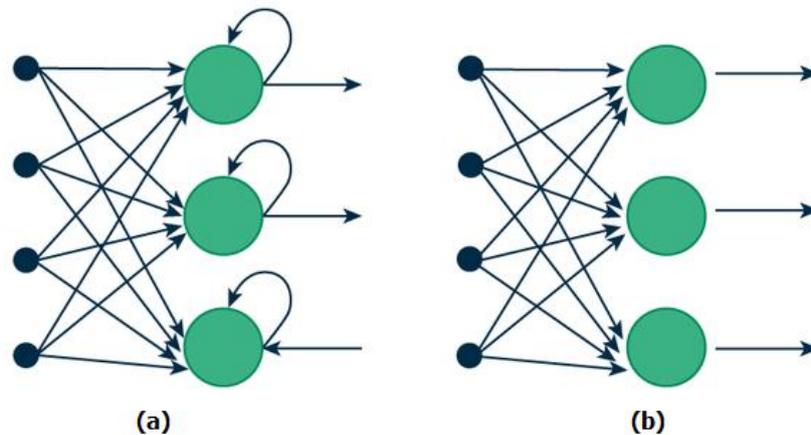
The hidden state acts as a memory, enabling the network to capture temporal dependencies and process sequential data, such as text or speech, where the order of data points is essential.

#### Applications of RNNs

Include speech recognition, where they process temporal audio data; NLP tasks like machine translation and sentiment analysis; time series prediction for forecasting data; video analysis for recognizing patterns in sequences of frames; and music composition, generating coherent sequences of notes.

#### Limitations of RNNs

Consist of the vanishing and exploding gradient problems, making it hard to learn long-term dependencies; difficulty in handling long-term dependencies; slow training times due to computational complexity; lack of parallelization due to sequential processing; and overfitting due to model complexity and large parameter sets. These limitations are often addressed by more advanced RNN variants like LSTMs and GRUs.



**Figure 4.5:** (a) RNNs with feedback connections, (b) FNNs process inputs independently without memory of past data.

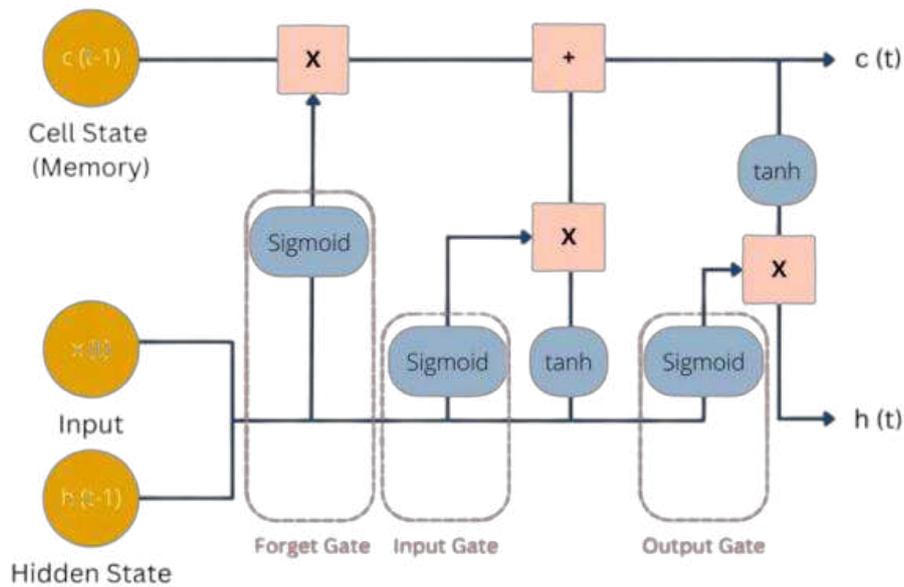
### 2.4.5 Long Short-Term Memory (LSTM) Networks

LSTMs are a specialized type of Recurrent Neural Network (RNN) designed to effectively capture and retain information over long sequences. Unlike conventional RNNs, they address the long-term dependency problem, allowing them to preserve relevant information for extended periods.

This capability is made possible through LSTM units shown in Figure 4.6, which incorporate key components such as the input, output, and forget gates. These gates regulate the flow of information within the network, determining which information should be stored, updated, or discarded. By dynamically managing memory, LSTMs enhance the network's ability to make accurate predictions based on historical data [138].

At each computational step, the current input  $x_t$ , the previous cell state  $C_{t-1}$ , and the previous hidden state  $h_{t-1}$  are processed to generate an updated cell state and hidden state. These values pass through three key gates that regulate the flow of information:

1. **Forget Gate:** This gate determines which past information should be retained or discarded. Both the previous hidden state and current input are passed through a sigmoid activation function, producing values between 0 and 1. A value close to 0 results in forgetting irrelevant information, while a value close to 1 retains important past information. The resulting values are then multiplied element-wise with the current cell state  $C_{t-1}$ , ensuring that only the necessary information is preserved.



**Figure 4.6:** LSTM units that forms the basic Long-Short Term Memory Networks architecture

2. **Input Gate:** This gate assesses the relevance of the new input to the current task. The current input is combined with the previous hidden state and passed through a sigmoid function to determine which values should be updated. Simultaneously, a candidate cell state is created using a tanh activation function to scale the new information. The filtered input information is then added to the cell state  $C_t$ , forming an updated long-term memory that will be used in the next time step.
3. **Output Gate:** This gate determines the final hidden state  $h_t$ , which serves as the LSTM's output. A sigmoid activation function selects the most relevant information to be passed to the output. The updated cell state undergoes a tanh activation, and the result is multiplied element-wise with the output filter, forming the new hidden state. Depending on the task, this hidden state could represent the next word in a sentence, a classification label, or another prediction relevant to the application.

This structured gating mechanism allows LSTMs to effectively capture long-term dependencies in sequential data, making them particularly useful for tasks such as natural language processing, speech recognition, and medical time-series analysis.

### 2.4.6 Convolution Neural Network (CNN)

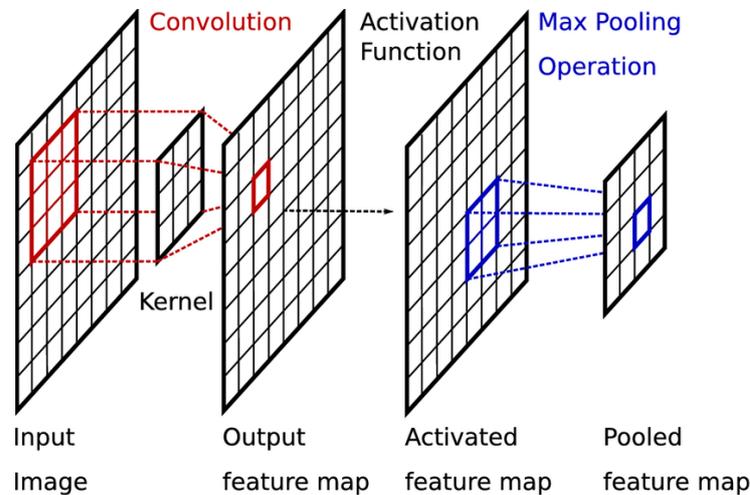
Convolutional Neural Networks (CNNs) are a specialized class of deep learning models designed to efficiently process grid-like data, particularly images. Unlike traditional fully connected neural networks, CNNs exploit spatial hierarchies using convolutional operations, which preserve local features while significantly reducing the number of trainable parameters. At the core of CNNs is the convolution operation, where a small matrix called a kernel (filter) illustrated in Figure 4.7 slide over the input image to extract features [139]. Mathematically, a 2D convolution is expressed as:

$$S(i,j) = (X * W)(i,j) = \sum_m \sum_n X(i+m, j+n)W(m,n)$$

where:

- $X(i,j)$  represents the input image or feature map,
- $W(m,n)$  is the kernel (filter), a small matrix that learns feature representations,
- $S(i,j)$  is the resulting feature map.

Each kernel detects specific patterns (e.g., edges, textures, shapes), and multiple kernels in deeper layers extract high-level hierarchical features [140]. This approach enables parameter sharing, reducing the computational cost compared to fully connected layers. To introduce non-linearity, CNNs apply an activation function, commonly ReLU which helps mitigate the vanishing gradient problem and accelerates training.

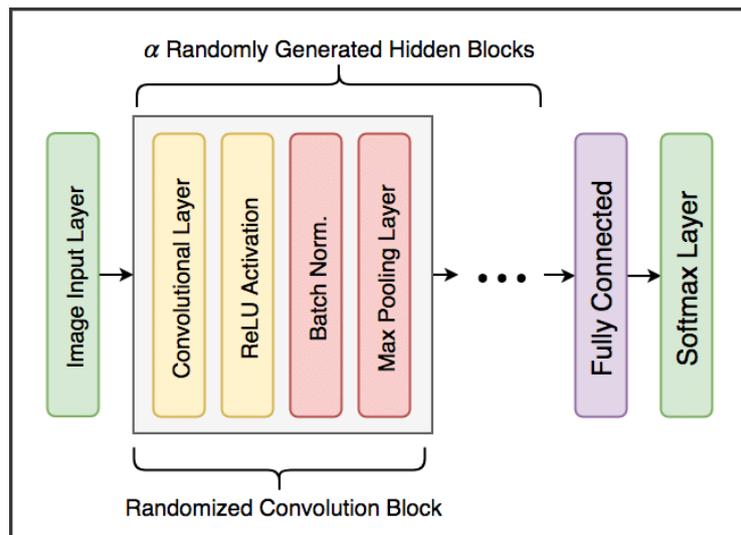


**Figure 4.7:** Convolution operation, the core concept of CNNs.

As described in Figure 4.8 CNN architectures process data through a series of hierarchical layers:

1. Convolutional Layers apply kernels to detect spatial features.
2. Pooling Layers (e.g., max pooling) downsample feature maps to reduce dimensionality.
3. Fully Connected Layers integrate high-level features for classification or regression.
4. Normalization and Regularization techniques (e.g., batch normalization, dropout) enhance model stability and prevent overfitting.

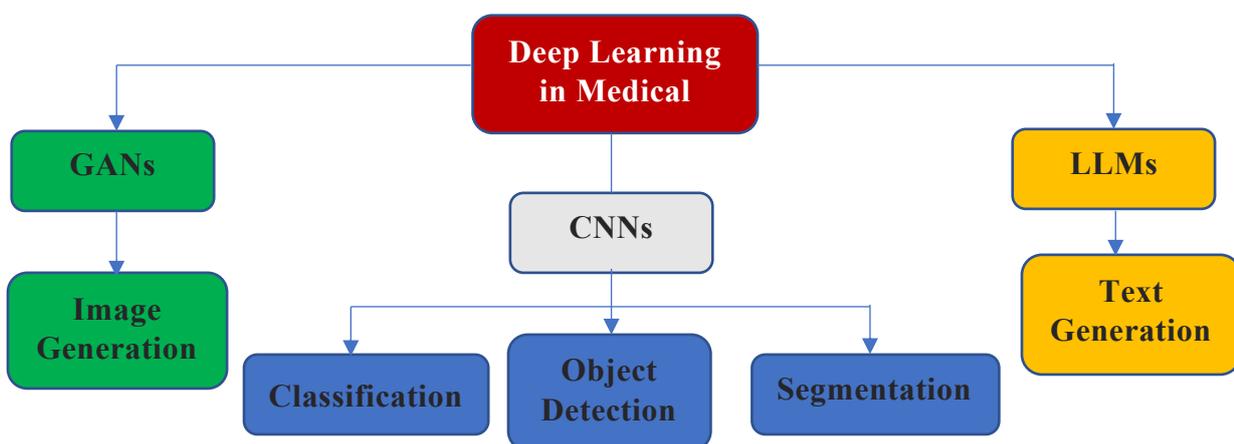
CNNs are particularly well-suited for medical imaging applications, such as classification of microcalcifications, tumor detection, and segmentation, as they can capture fine-grained patterns in high-resolution medical images [140].



**Figure 4.8:** General structure of an individual CNN network with  $\alpha$  convolution blocks.

## 2.5 Deep learning in medical imaging and breast cancer

As highlighted in Figure 4.9 deep learning models in medical imaging are designed to process images for various tasks, such as detecting abnormalities, enhancing image quality by reducing noise and improving spatial resolution, and supporting additional functions like automated report generation and information retrieval. Additionally, data augmentation techniques such as rotation, flipping, scaling, and contrast adjustments are commonly employed to artificially expand training datasets, improving model robustness and generalization [141]. In the following sections, we will briefly review key deep learning techniques applied in medical imaging before delving into their specific applications in breast cancer imaging.



**Figure 4.9:** Deep learning application in medical imaging [141].

Numerous commercial products have been introduced and received regulatory approval from the U.S. Food and Drug Administration (FDA), making them officially authorized for clinical use. Additionally, numerous retrospective and prospective studies have been carried out to evaluate their performance. Table 4.1 summarized AI-based tools deployed for breast cancer detection and diagnosis.

<b>Tool</b>	<b>Vendor</b>	<b>Country</b>	<b>Modality</b>	<b>Techniques</b>
<b>cmAssist®</b>	CureMetrix Inc.	United States	MG	Deep artificial neural networks (dANNs) trained on radiologist-annotated mammograms.
<b>Genius AI Detection</b>	Hologic Inc.	United States	MG + DBT	CNN-based classification integrated with tomosynthesis.
<b>INSIGHT MMG</b>	Lunit Inc.	South Korea	MG	Fourfold ConvNeXt-small network; winner of RSNA 2022 challenge.
<b>MammoScreen® 2.0</b>	Therapixel SA	France	MG + DBT	Synthetic DBT image generation for streamlined workflows.
<b>ProFound AI®</b>	iCAD Inc.	United States	MG + DBT	CNN-based radiomics for microcalcifications; automated lesion localization.
<b>Saige-Dx</b>	DeepHealth Inc.	United States	MG	Deep learning classifiers for triaging mammograms.
<b>Transpara®</b>	ScreenPoint Medical B.V.	Netherlands	MG + DBT	CNN-based risk scoring with multi-view analysis.

**Table 4.1:** Principal commercialized AI-based tools for breast cancer [141].

In what follow, an overview of multiple retrospective and prospective studies on breast cancer detection in mammography that evaluated AI-based tools in the table below and other techniques.

Watanabe et al. [142]. Demonstrated a retrospective analysis of the AI-CAD software cmAssist® on 122 patients (90 false-negative mammograms) showed an 11% increase in radiologists' cancer detection rate (CDR). The AI improved sensitivity but was tested in a cancer-enriched cohort, potentially inflating results. Funded by the vendor (CureMetrix), the study raised concerns about bias and generalizability.

chaffter et al. [143]. Published the result of a public challenge involved 31 teams developing AI models for mammography using 85,000 U.S. and 68,000 Swedish screening exams. The top model achieved an AUC of 0.858 (U.S.) and 0.903 (Sweden). While

standalone AI underperformed radiologists (specificity = 66.2% vs. 90.5%), combining AI with radiologists improved AUC to 0.942.

Kim et al. [144]. A retrospective study tested INSIGHT MMG on 320 mammograms (160 cancers). The standalone AI achieved an AUC of 0.940, outperforming unassisted radiologists (AUC = 0.810). AI-assisted radiologists reached an AUC of 0.881. Strengths included detection of small tumors (<10 mm), but the cancer-enriched dataset limited real-world applicability.

Dembrower et al. [145]. These studies evaluated INSIGHT MMG in retrospective (2020) and prospective (2023) settings. The AI triaged mammograms, reducing radiologist workload by 50% without missing cancers. In a prospective trial with 55,000+ women, AI-assisted single reading matched double radiologist reading in cancer detection. Limitations included vendor-specific hardware and short follow-up.

Ng et al. [146]. A prospective trial tested INSIGHT MMG as an adjunct to double reading, detecting 0.7–1.6 additional cancers per 1,000 cases with minimal unnecessary recalls. The AI improved early detection of invasive cancers but faced criticism for short follow-up (2–9 months) and single-institution data.

Romero-Martín et al. [147]. This retrospective study assessed Transpara® on 15,999 MG/DBT exams. The AI achieved AUCs of 0.93 (MG) and 0.94 (DBT), demonstrating non-inferiority to radiologists. However, DBT recall rates rose by 12.3%, increasing false positives. The study emphasized workflow efficiency but lacked clinical outcome data.

Zheng et al. [148]. A prospective multicenter study used a RefineNet + Xception model for contrast-enhanced mammography (CEM). The AI automated lesion segmentation (DSC = 0.837) and classification (AUC = 0.891), proving effective for single-mass lesions. Limitations included a focus on Chinese women, limiting generalizability.

Beuque et al. [149]. This study combined DL segmentation + handcrafted radiomics on CEM images, achieving an AUC of 0.95 for lesion classification. Automated segmentations outperformed manual ones, but the small sample size (retrospective design) and single-center data reduced external validity.

Despite significant advancements in AI-based breast cancer detection, several critical limitations and challenges must still be overcome for successful clinical integration. One major issue is dataset bias, as many studies use cancer-enriched or institution-specific datasets, limiting AI's generalizability to diverse populations. Additionally, AI models often lack the ability to analyze multiple mammographic views or track temporal changes, both essential for detecting subtle or evolving malignancies. Performance inconsistencies also arise based on breast density, with AI struggling in dense tissue where cancers are harder to distinguish. Furthermore, most research has been retrospective, conducted under controlled conditions that do not fully replicate real-world clinical workflows, leading to concerns about overfitting and observer bias. Ethical and regulatory challenges add complexity, including liability concerns, data privacy issues, and the need for rigorous validation before AI tools can be widely adopted. The integration of AI into radiology workflows requires significant infrastructure adaptations, such as PACS compatibility and

clinician training, to ensure seamless implementation. Lastly, while AI has demonstrated the ability to enhance radiologists' performance, no standalone system has yet surpassed human expertise, reinforcing the notion that AI should serve as an assistant rather than a replacement. Addressing these challenges through large-scale prospective trials, improved dataset diversity, and optimized AI-human collaboration will be essential for the widespread adoption of AI in breast cancer screening and diagnosis. High false and false negative rates is also a major problem faced in most of proposed solutions [156].

### **3. Ensemble Deep Learning**

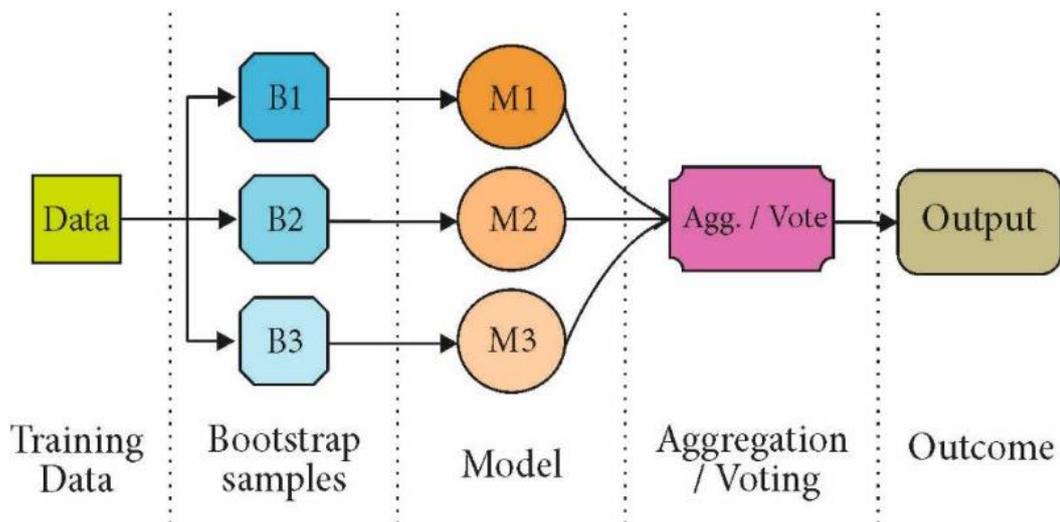
In machine learning, two approaches have demonstrated superior performance compared to traditional algorithms: ensemble learning and deep learning [153]. Ensemble learning within deep learning refers to a methodological framework that integrates multiple individual models to enhance predictive accuracy and robustness beyond what a single model could achieve independently. This approach aims to capitalize on the strengths of diverse models while mitigating their respective limitations, thereby improving generalization and reliability in complex predictive tasks.

Various ensemble learning techniques differ in how they train and combine distinct baseline models. The most commonly employed methods include averaging, bagging, stacking, and boosting. Each technique offers unique advantages and is suited for different types of data and modeling challenges [154].

#### **3.1 Bagging (Bootstrap Aggregating)**

Bagging is an ensemble learning technique designed to reduce variance and enhance model stability by training multiple instances of the same base model on different random subsets of the dataset. These subsets are created through bootstrap sampling, a process in which samples are drawn with replacement, ensuring that each training subset varies slightly from the original dataset. Once trained, the models produce independent predictions, which are then aggregated either by majority voting in classification tasks or averaging in regression tasks. The most well-known algorithm utilizing bagging is Random Forest, which constructs multiple decision trees and combines their outputs to produce a more robust and accurate prediction than a single tree. Bagging is particularly effective in reducing overfitting in high-variance models like decision trees, improving generalization to unseen data [150].

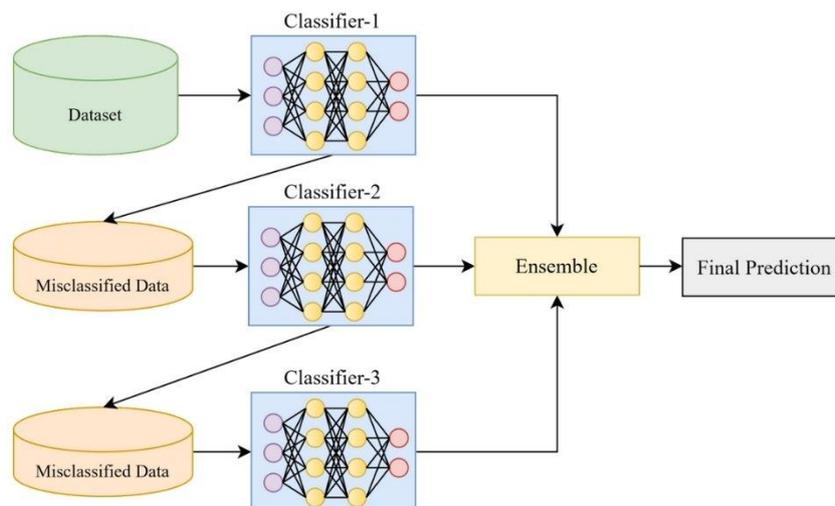
In deep learning bagging refers to training multiple deep neural networks independently on different subsets of the training data and then combines their outputs as shown in Figure 4.10. Each model is trained on a randomly sampled (with replacement) subset of the data, ensuring diversity among the models. The final prediction is obtained through majority voting (classification) or averaging (regression). In deep learning, bagging is particularly useful in reducing the variance of high-capacity models like deep convolutional or transformer networks. Deep Ensembles, a popular approach in uncertainty estimation, utilizes bagging to train multiple deep neural networks independently and aggregate their predictions, improving both accuracy and reliability in tasks like medical diagnosis and autonomous driving.



**Figure 4.10:** Bagging ensemble learning approach.

### 3.2 Boosting

Boosting is an advanced ensemble method designed to reduce bias by sequentially training models in a way that prioritizes correcting previous errors. Unlike bagging, where models are trained independently, boosting follows a stage-wise approach, where each new model is trained to improve upon the weaknesses of the preceding models. During training, misclassified instances receive higher weights, forcing the model to focus on the most challenging examples. The final prediction is obtained by combining the outputs of all models, often through weighted voting or summation. Prominent boosting algorithms include AdaBoost, which assigns adaptive weights to misclassified samples, and Gradient Boosting (GBM, XGBoost, LightGBM, CatBoost), which optimizes a loss function by sequentially training models on residual errors.



**Figure 4.11:** Boosting Approach in Ensemble Learning.

As illustrated in Figure 4.11 boosting in deep learning differs from its classical implementation due to the nature of neural networks, which require extensive training. Instead of sequentially training weak learners, boosting in deep learning often involves

adaptive weighting of data samples or feature representations. One effective approach is Gradient Boosted Neural Networks (GBNN), where multiple neural networks are trained in sequence, with each subsequent model focusing on minimizing the errors of the previous ones. Another adaptation is Boosted Convolutional Neural Networks (BoostCNNs), where smaller CNN models learn residual patterns missed by previous models, improving accuracy on challenging image classification tasks. While boosting can improve performance, it requires careful regularization to prevent overfitting in deep networks. Also, boosting is particularly effective in scenarios requiring high predictive accuracy but may suffer from overfitting if not properly regularized.

### 3.3 Stacking (Stacked Generalization)

Stacking, or stacked generalization, in deep learning involves training multiple heterogeneous deep models and using a meta-model to combine their predictions. Unlike bagging and boosting, which typically use homogeneous networks, stacking leverages different architectures as illustrated in Figure 4.12, such as CNNs, recurrent neural networks (RNNs), and transformer models, to extract complementary features from the data. A common deep learning stacking strategy involves training base models separately, then using a fully connected neural network (meta-learner) to process their outputs and make the final prediction. Stacking is widely used in image recognition, speech processing, and multimodal learning, where combining different feature representations leads to superior accuracy and generalization. However, stacking is computationally expensive and requires careful tuning to avoid overfitting.

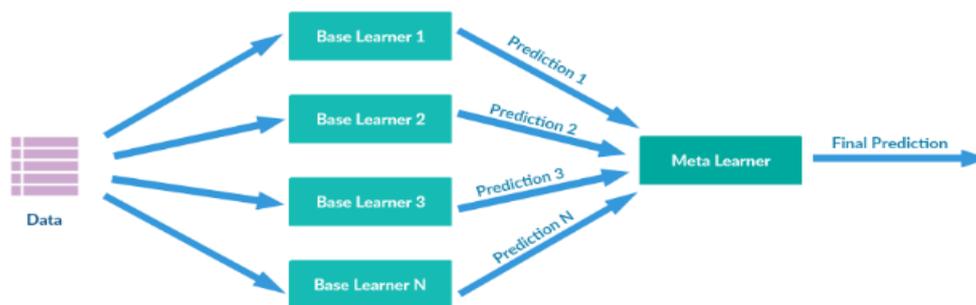


Figure 4.12: Architecture of stacking ensemble model using meta learner.

### 3.4 Voting and Averaging

Voting and averaging are ensemble techniques that aggregate predictions from multiple models to enhance stability and performance. In hard voting, the final classification decision is made based on the majority prediction among all models. In soft voting, the probability distributions of each model's predictions are averaged, and the class with the highest probability is selected. For regression tasks, simple averaging is used, where the final output is computed as the mean of all individual model predictions. Voting ensembles are particularly useful when combining diverse models, as they help smooth out

individual model biases. While easy to implement, voting ensembles may not always yield significant improvements unless the base models are sufficiently diverse.

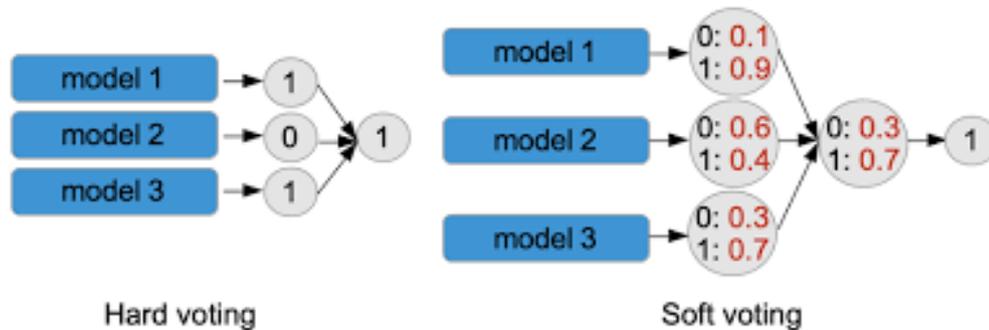


Figure 4.13: Hard and soft voting

### 3.5 Blending

Blending is a variation of stacking that simplifies the training process of the meta-model by using a holdout validation set instead of cross-validation predictions. In this approach, base models are trained on the training data, and their predictions on a separate validation dataset serve as inputs to the meta-model, which learns the best way to combine them. The advantage of blending is its ease of implementation, as it avoids the complexity of cross-validation required in stacking. However, blending can suffer from overfitting if the validation set is too small or not representative of the full dataset. This technique is widely used in Kaggle competitions, where time efficiency is critical, but it is less common in real-world applications due to its reliance on a single validation set.

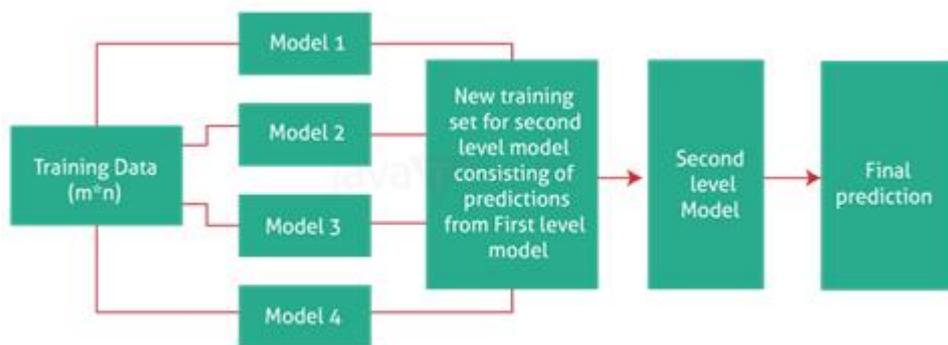


Figure 4.14: Blending strategy used ensemble learning.

### 3.6 Knowledge Distillation as an Implicit Ensemble Learning Method

Knowledge distillation, though not a conventional ensemble method, can be considered a form of implicit ensembling. In this technique, a large ensemble (teacher model) is trained first, and its knowledge is transferred to a smaller, more efficient student model. The student network learns not only from the original data labels but also from the soft probabilities of the teacher's predictions, capturing richer information. This method is

widely used in efficient deep learning, where computational constraints require lighter models with ensemble-level performance, such as in mobile and edge AI applications.

Each ensemble learning technique has its strengths and limitations, making them suitable for different types of tasks. Bagging is ideal for reducing variance and improving stability, while boosting excels at reducing bias and enhancing predictive accuracy. Stacking provides a powerful mechanism for leveraging multiple diverse models, though it requires careful tuning. Voting and averaging offer a straightforward way to improve performance through model aggregation, and blending provides a practical alternative to stacking. The selection of an ensemble technique depends on factors such as dataset size, computational constraints, and the need for interpretability versus accuracy.

## **4. Classification of Microcalcifications using Ensemble Learning**

Ensemble learning strategies in deep learning integrate multiple models to enhance predictive performance, mitigate overfitting, and optimize classification accuracy [176]. By utilizing diverse model architectures, subsets of training data, or varying learning approaches, ensemble models are capable of capturing a broader range of patterns and generalizing more effectively than a single model. Several techniques are commonly employed to implement ensemble models, with bagging, boosting, voting, and stacking being among the most prevalent. Despite their widespread application, relatively few studies have explored the use of stacking models for the classification of microcalcifications.

This section introduces an innovative ensemble framework that combines several pre-trained convolutional neural networks to improve the classification of microcalcifications, tiny calcium deposits that play a key role in early breast cancer detection. To better understand and enhance their performance, we take a close look at three well-known architectures: ResNet-50, DenseNet-121, and EfficientNet-B0. Through evaluation, we explore their strengths and limitations, not just in theory but also in terms of how they can be effectively used in real clinical settings. Our goal is to identify models that are not only accurate but also practical for deployment in automatic diagnostic. In addition, we experiment with different ensemble learning strategies to boost the performance of these base models, aiming to deliver more reliable and precise classifications. By bridging the gap between cutting-edge deep learning techniques and real-world healthcare needs, this work contributes to improving clinical decision-making and patient outcomes [164].

### **4.1 Related works**

The diagnosis of breast cancer, particularly the detection and classification of microcalcifications, has significantly advanced in the past decade due to the integration of machine learning in automatic diagnosis frameworks. Ongoing research continues to enhance accuracy, reliability, model architecture, data augmentation, and interpretability, marking a new era in diagnostic imaging.

Cai et al. [156] introduced a computer-aided detection and diagnosis (CADE/CADx) system designed for the identification of microcalcification clusters (MCCs) in mammograms. Their approach leveraged a deep convolutional neural network (DCNN) to replace traditional manual feature extraction, incorporating neutrosophic boosting during training. The model achieved a sensitivity of 90% at 0.14 false positives per image for MCC detection and demonstrated an area under the receiver operating characteristic curve (AUC) of 0.945 on the validation dataset and 0.933 on the test dataset using the INbreast dataset.

Kang et al. [157] explored the use of five pre-trained DCNN architectures, ResNet-101, Xception, Inception-v3, InceptionResNet-v2, and DenseNet-201, alongside an ensemble model for classification. Using 1,579 mammographic images, their best model attained an accuracy of 81.54%, a specificity of 91.41%, and a sensitivity of 82.47%.

Gerbasi et al. [158] proposed DeepMiCa, a three-step system designed for microcalcification detection. The first step involved preprocessing raw scan data, followed by automated patch-based semantic segmentation via a UNet-based network equipped with a custom loss function tailored for extremely small lesions. Finally, lesion classification was performed using deep transfer learning. The approach was evaluated on the CBIS-DDSM and INbreast datasets, achieving an AUC of 0.89.

Tsai et al. [159] developed a detection framework based on VGG16, Mask R-CNN, and Inception V3, reporting respective precision values of 93.63%, 99.76%, and 88.89%. Similarly, Teoh et al. [160] proposed a framework integrating morphological operations and Otsu segmentation, followed by transfer learning with ResNet-50 and ensemble optimization utilizing AlexNet, GoogLeNet, VGG16, and ResNet-50. Their model achieved an average confidence level of 0.9305 in classification.

Du et al. [161] introduced a novel approach by transforming the characterization of morphology and distribution of microcalcifications into a node and graph classification problem, employing a multi-task deep graph convolutional network (GCN). Trained and validated on the DDSM dataset, the model achieved an AUC of 0.87. Yurdusev et al. [162] enhanced microcalcification detection by applying a difference filter to amplify relevant regions, subsequently utilizing Faster R-CNN and YOLOv4 deep learning models on 500 mammograms from the DDSM dataset, attaining an accuracy of 97%.

Singh et al. [163] developed a deep learning-based classification model to categorize microcalcifications into benign, malignant, and benign without callback categories. Using the CBIS-DDSM dataset, they employed a pretrained InceptionResNetV2 model for feature extraction and experimented with four optimizers (ADAM, ADAGrad, ADADelta, RMSProp). Their model achieved a sensitivity of 97%, specificity of 80%, accuracy of 94%, and an AUC of 96%.

A review of the existing literature highlights a clear tendency in the direction of the integration of multiple models and hybrid approaches, combining various deep learning techniques to enhance diagnostic performance. The transition from traditional machine learning to deep learning, and subsequently to ensemble and hybrid models, has led to

substantial improvements in microcalcification detection and diagnosis. Despite these advancements, challenges remain, particularly in reducing false positive rates, improving the interpretability of deep learning models, and ensuring consistent performance across diverse imaging conditions. Addressing these limitations will be crucial for the successful clinical implementation of AI-driven diagnostic tools in breast cancer screening.

## 4.2 Data description

The Digital Database for Screening Mammography (DDSM) is one of the largest publicly available mammographic datasets, containing over 10,000 images collected from various screening programs. It has been widely used for breast cancer research, particularly in early machine learning models and radiology studies. The dataset includes both benign and malignant cases, with annotations for masses and microcalcifications.

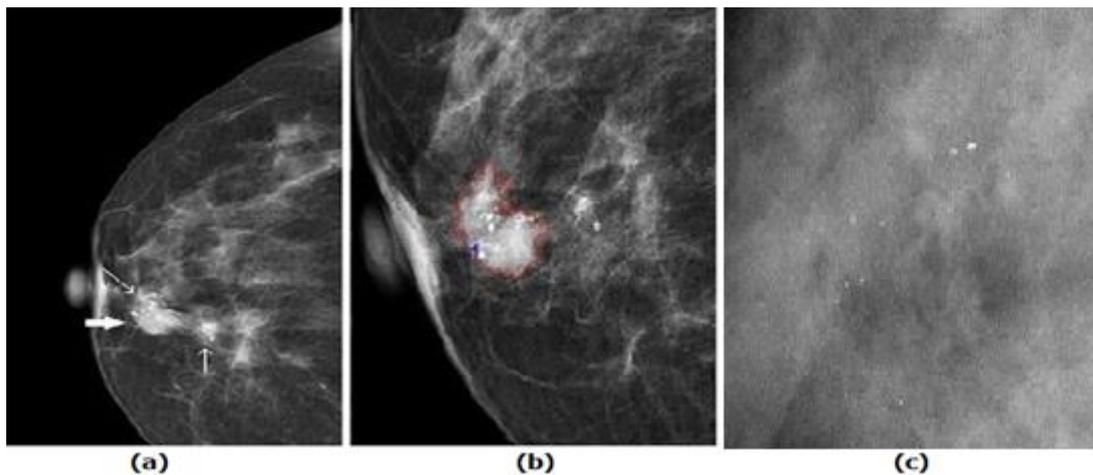
Despite its extensive size and historical significance, DDSM presents several challenges that make it less ideal for deep learning applications. First, the images are stored in lossless JPEG (LJPEG) format, which is not natively supported by modern deep learning frameworks. Researchers must convert these images to more commonly used formats such as PNG or DICOM, requiring additional preprocessing steps.

Another major challenge is the quality of annotations. The dataset includes manually annotated regions of interest (ROIs), but some segmentations are inconsistent or imprecise, which can affect model performance. Additionally, since the images were collected from different sources, there is variation in image quality and contrast, making it difficult to standardize data for deep learning applications.

In spite of these challenges, DDSM remains a valuable resource for traditional machine learning and radiology-focused studies, particularly for researchers willing to invest time in data cleaning and preprocessing.

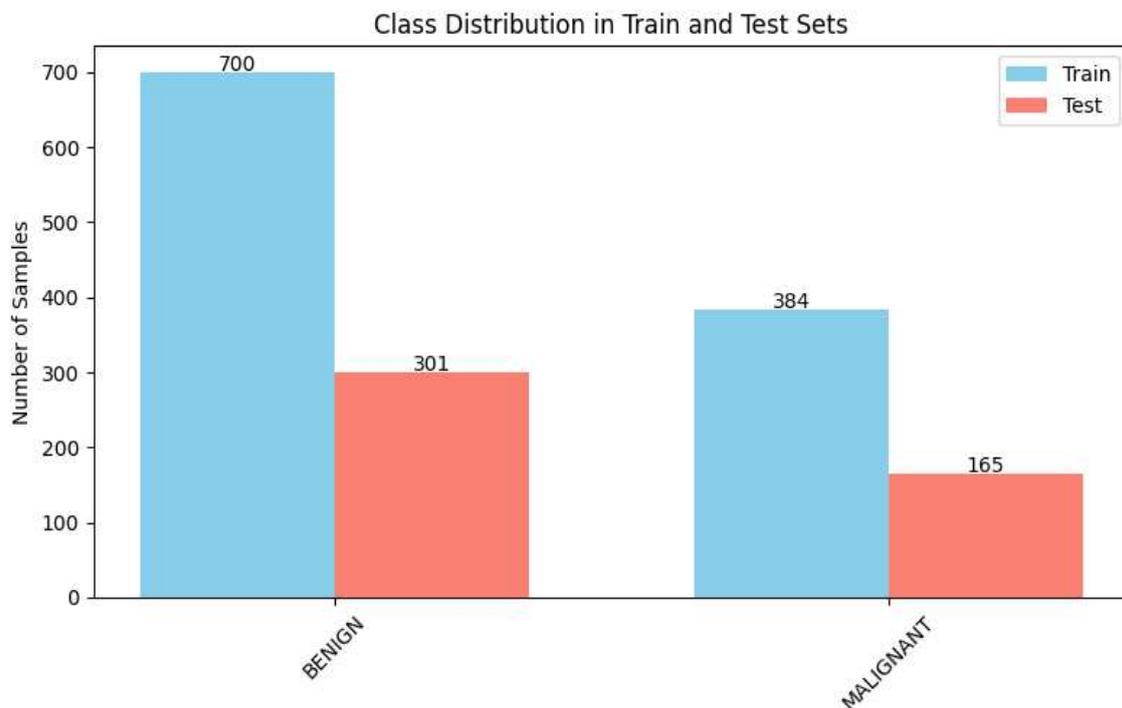
To address the challenges of DDSM, researchers at The Cancer Imaging Archive (TCIA) [108] developed the Curated Breast Imaging Subset of DDSM (CBIS-DDSM). This dataset is a refined version of DDSM, specifically designed to facilitate deep learning applications in breast cancer detection [165].

One of the most significant improvements in CBIS-DDSM is the adoption of the DICOM (Digital Imaging and Communications in Medicine) format. Unlike LJPEG, DICOM is a widely accepted medical imaging standard, making it directly compatible with modern deep learning frameworks and radiology software. This eliminates the need for complex image format conversions, allowing researchers to focus on model development. Additionally, CBIS-DDSM provides improved ROI segmentations as illustrated in Figure 4.15. The original DDSM annotations were reviewed and refined for better accuracy, ensuring that deep learning models can learn from precise tumor and microcalcification boundaries. This reduces errors caused by noisy labels and improves model training efficiency [116].



**Figure 4.15:** (a) Suspicious abnormality detected in image (a), after segmentation in image (b), ROI image in image (c) is extracted to be used for characterization by deep learning model.

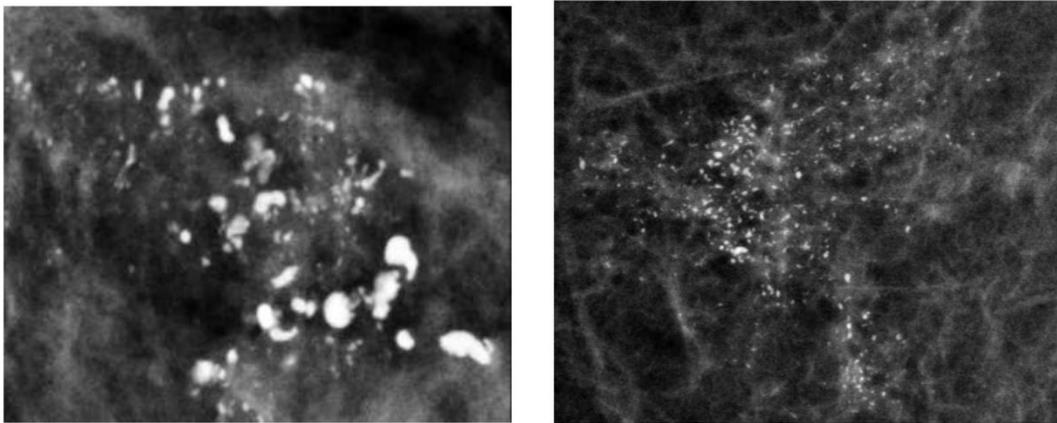
Another key advantage of CBIS-DDSM is its easier accessibility and structured format. The dataset has been preprocessed and standardized, removing many of the inconsistencies found in DDSM. This makes it ideal for quick deployment in deep learning experiments, allowing researchers to focus on model architecture rather than extensive data preparation. Due to these enhancements, CBIS-DDSM is considered the preferred choice for deep learning research in mammographic analysis, particularly for the detection of masses and microcalcifications.



**Figure 4.16:** Class distribution in training and test dataset of CBIS-DDSM.

Experiments conducted in our contributions in this chapter, utilized 1550 images: 1085 (70%) for training and 465 (30%) for the test phase, as detailed in Figure. 4.16. All extracted images from CBIS-DDSM are either cranial-caudal (CC) or mediolateral oblique (MLO) mammograms. Figure 4.17, shows image samples taken from CBIS-DDSM dataset.

The choice between DDSM and CBIS-DDSM depends on the specific research objective. For researchers focused on deep learning and AI-based breast cancer detection, CBIS-DDSM is the better choice due to its cleaner annotations, standardized format, and easier accessibility. It allows for a faster and more efficient model development process, reducing the need for extensive data preprocessing.



**Figure 4.17:** Image samples taken from CBIS-DDSM dataset.

### 4.3 Pre-processing

Data preprocessing plays a fundamental role in developing highly accurate models for the classification of microcalcifications in mammographic images. However, their small size, subtle contrast variations, and diverse morphological appearances across different patients pose significant challenges for detection and classification. To enhance model performance and improve the reliability of diagnostic outcomes, effective preprocessing techniques are essential. These techniques aim to refine image quality, suppress noise, and standardize data, ultimately facilitating the extraction of relevant features for deep learning algorithms. Our approach focuses on three key preprocessing techniques; noise reduction, normalization & standardization, and contrast enhancement, to optimize mammographic image quality and improve the accuracy of microcalcification detection. These preprocessing steps are essential for ensuring that deep learning models can effectively learn and distinguish microcalcifications from surrounding breast tissue, thereby enhancing diagnostic performance.

#### 4.3.1 Noise Reduction

Mammographic images often contain unwanted background noise, texture variations, and granular artifacts, which can obscure microcalcifications and reduce the sensitivity of detection algorithms. To address this issue, we employ Non-Local Means (NLM) denoising [167], a technique that selectively reduces noise while preserving critical image details. By

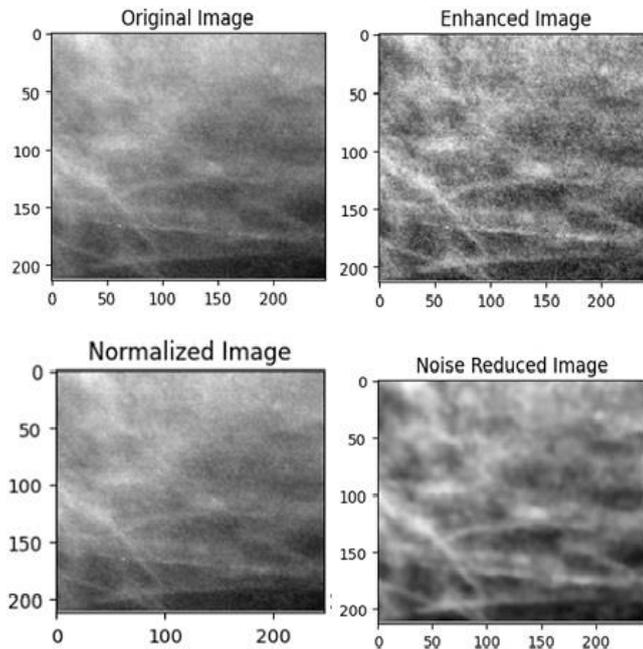
eliminating irrelevant patterns, NLM denoising enhances the clarity of microcalcifications, allowing the model to focus on diagnostically relevant features.

### 4.3.2 Normalization and Standardization

Variability in mammographic images arises from differences in imaging equipment, acquisition protocols, and patient-specific anatomical factors, leading to inconsistencies in brightness, contrast, and intensity levels. To ensure uniformity across images, we apply normalization and standardization techniques. Normalization scales pixel intensities to a fixed range (e.g., 0 to 1), while standardization adjusts the data distribution to have a mean of 0 and a standard deviation of 1. These transformations help reduce data variability, making it easier for the deep learning model to generalize across different datasets and imaging conditions.

### 4.3.3 Contrast Enhancement

Microcalcifications often exhibit low contrast against surrounding breast tissue, making them difficult to detect. To enhance their visibility, we integrate Histogram Equalization (HE) and Contrast-Limited Adaptive Histogram Equalization (CLAHE) [168] into our preprocessing pipeline. HE redistributes intensity values to enhance global contrast, while CLAHE applies localized contrast adjustments to improve the visibility of small, subtle structures without over-enhancing other regions. This step ensures that microcalcifications are more distinguishable, benefiting both radiologists and AI-based detection models.



**Figure 4.18:** Different preprocessing techniques applied on mammograms.

Each of the preprocessing steps outlined above plays a crucial role in refining the dataset, ensuring greater consistency and accuracy, which in turn improves model performance and facilitates automated diagnostic processes. Figure 4.16 illustrates the original image alongside the various preprocessing tasks that have been applied. The

sequence in which these tasks are performed is critical for optimizing the preprocessing outcomes. Specifically, noise reduction is applied first, followed by normalization, with image enhancement serving as the final step. This strategic order maximizes the effectiveness of the preprocessing pipeline.

## **4.4 Overview of CNN architectures employed in our ensemble learning strategy**

In this ensemble approach for breast cancer classification, we combine three powerful convolutional neural networks: ResNet-50, DenseNet-121, and EfficientNetB0. Each model brings its own unique strengths to the table. ResNet-50 uses skip connections that help it avoid common training issues like vanishing gradients, making it easier to train deeper networks while still being lightweight and efficient great for medical imaging tasks. DenseNet-121 stands out with its dense connections, where each layer gets input from all previous ones. This design encourages feature reuse and helps the model learn more effectively, especially when data is limited. EfficientNetB0 takes a different approach, using a smart scaling method to balance depth, width, and resolution, resulting in a compact yet highly accurate model. It's the product of neural architecture search, which means it's been fine-tuned for both performance and efficiency. By combining these models using different ensemble strategies like stacking or boosting, the system takes advantage of their complementary strengths to make more robust and accurate predictions.

### **4.4.1 ResNet-50**

ResNet-50 [175] is a convolutional neural network (CNN) architecture that belongs to the ResNet (Residual Networks) family, consisting of 50 layers. It was specifically designed to address the challenges associated with training very deep networks. Developed by Microsoft Research Asia, ResNet-50 has gained widespread recognition for its depth and computational efficiency, particularly in image classification tasks. It leverages multiple residual blocks, as illustrated in Figure 4.19, to achieve robust performance.

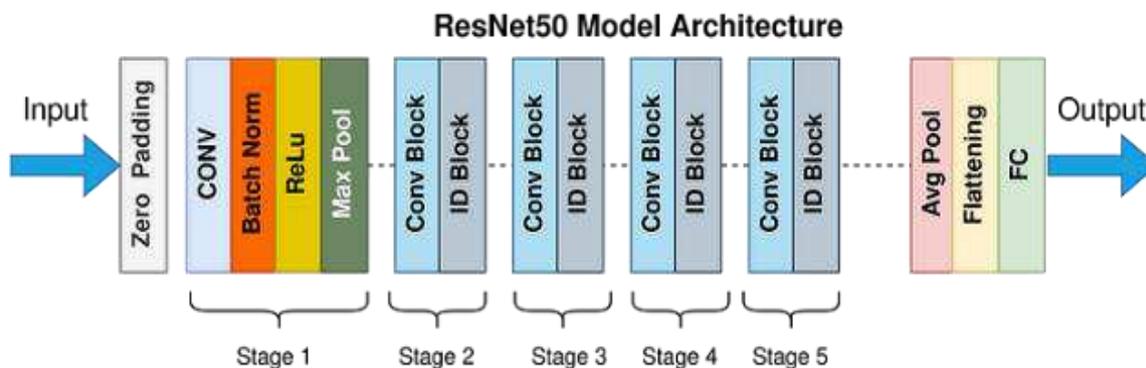
The key innovation in ResNet-50 is the concept of residual learning, which is the foundation of the residual blocks. In traditional deep networks, each layer learns a mapping from the input to the output. However, as the network becomes deeper, the gradients during backpropagation can diminish, making it difficult for the model to learn effectively. The residual learning concept addresses this issue by introducing skip connections. As highlighted in Figure 4.18 these skip connections add the input of a residual block directly to its output, allowing the network to learn the residual mapping (i.e., the difference between the input and output) rather than the full transformation. This simple yet powerful modification enables the network to preserve important information, alleviates the vanishing gradient problem, and facilitates the training of much deeper networks.



**Figure 4.19:** Concept of Residual blocks in ResNet networks.

Residual blocks typically consist of a series of convolutional layers, followed by batch normalization and ReLU activation functions. By learning the residuals rather than the entire mapping, the network is able to focus on the incremental changes needed to improve the prediction. This mechanism allows for the successful training of very deep networks without encountering performance degradation, which was a common issue in earlier architectures.

ResNet models come in various configurations, with alternative versions such as ResNet-18 and ResNet-32, each differing in depth and complexity. Since its introduction in 2015, ResNet-50 described in Figure 4.20 has played a pivotal role in advancing the field of image classification, influencing subsequent developments in deep learning architectures and demonstrating continued relevance in both academic and practical applications.



**Figure 4.20:** ResNet-50 architecture

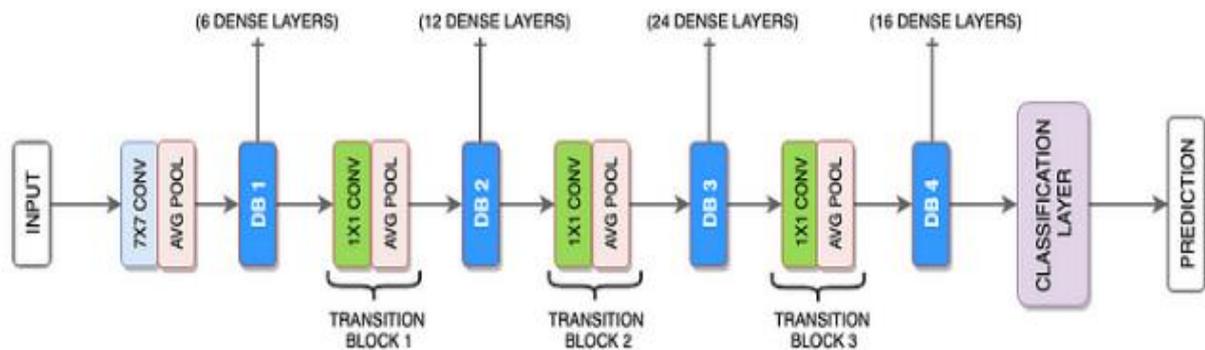
#### 4.4.2 DenseNet-121

DenseNet [169], short for Densely Connected Convolutional Networks, is a convolutional neural network (CNN) architecture introduced in 2016 by Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. It has garnered significant attention for its outstanding performance in image classification tasks and its innovative approach to information flow within deep networks. What sets DenseNet apart from conventional CNN architectures is its unique connectivity pattern, in which each layer is directly connected to every other subsequent layer in a feedforward manner. This densely connected structure ensures maximum information flow between layers throughout the network. As illustrated in Figure 4.21, DenseNet is characterized by two fundamental architectural innovations.

First, the use of dense blocks significantly enhances both feature propagation and gradient flow. In each dense block, every layer receives as input the concatenated outputs of all preceding layers. This strategy not only promotes feature reuse but also alleviates the vanishing gradient problem, which commonly hampers the training of deep networks. As a result, DenseNet can be trained with fewer parameters while maintaining or even improving representational power.

Second, DenseNet incorporates bottleneck layers and transition layers to improve computational efficiency and control model complexity. Bottleneck layers, typically consisting of  $1 \times 1$  convolutions followed by  $3 \times 3$  convolutions, serve to reduce the number of input feature maps, thereby decreasing the computational cost. Transition layers, which include convolution and pooling operations, are used between dense blocks to manage the dimensionality of the feature maps and further optimize the model's depth and performance.

This architectural design allows DenseNet to achieve a high level of parameter efficiency and accuracy compared to other deep CNN models, making it particularly effective for image classification and related computer vision tasks. Its dense connectivity not only improves gradient flow and convergence but also enables the network to learn more compact and robust representations of the input data.



**Figure 4.21:** DenseNet-121 architecture

### 4.4.3 EfficientNet-b0

EfficientNet [162] represents a family of convolutional neural networks (CNNs) specifically designed to deliver high predictive accuracy while maintaining computational efficiency. As illustrated in Figure 4.22, EfficientNet distinguishes itself through the use of a compound scaling technique, which uniformly scales three critical dimensions of the network, depth (number of layers), width (number of channels per layer), and input resolution (size of the input image). This balanced scaling strategy enables the model to achieve superior performance with significantly fewer parameters and lower computational costs compared to traditional CNN architectures.

A fundamental component of EfficientNet is the stem layer, which serves as the network's entry point for raw image data. The stem consists of an initial convolutional block, typically a  $3 \times 3$  convolution with a stride of 2, followed by batch normalization and a non-linear activation function (commonly Swish in EfficientNet). The purpose of the

stem is to transform high-dimensional input images into compact, low-level feature maps that can be efficiently processed by deeper layers of the network. This transformation includes reducing the spatial resolution while increasing the number of feature channels, as result facilitating the extraction of essential visual patterns such as edges, corners, and textures. The design of the stem is critical because it sets the foundation for all subsequent computations, any information loss at this stage could hinder the model’s ability to learn complex representations.

Following the stem, EfficientNet employs a series of Mobile Inverted Bottleneck Convolution (MBConv) blocks, which are the core computational units of the architecture. These blocks are adapted from the MobileNetV2 architecture and are specifically designed to maximize efficiency without compromising representational power. An MBConv block comprises several stages:

1. **Expansion Phase:** A  $1\times 1$  pointwise convolution is used to expand the input feature map to a higher dimensional space, allowing the model to learn richer representations.
2. **Depthwise Convolution:** A lightweight convolution is applied to each channel independently using a  $3\times 3$  or  $5\times 5$  kernel. This reduces computational complexity compared to standard convolutions, as it avoids inter-channel mixing at this stage.
3. **Squeeze-and-Excitation (SE) Module** (optional but common in EfficientNet): This module adaptively recalibrates channel-wise feature responses by modeling interdependencies between channels. This technique is called attention mechanism and helps the network focus on the most informative features.
4. **Projection Phase:** Another  $1\times 1$  convolution projects the expanded features back to a lower-dimensional space, effectively acting as a bottleneck.
5. **Residual (Skip) Connection:** If the input and output dimensions match, a skip connection is added, allowing the input to bypass the block. This facilitates gradient flow during backpropagation and improves training stability.

The “inverted bottleneck” terminology branches from the architecture’s design, which first expands the number of channels (in contrast to traditional bottleneck designs that compress first), processes the data with a depth wise convolution, and then projects it back to a lower dimension. This inverted structure enhances information flow and computational efficiency.

By stacking these MBConv blocks in varying numbers and configurations depending on the specific EfficientNet variant (e.g., B0 to B7), the network can achieve a desirable trade-off between accuracy and resource usage. The architecture is the result of neural architecture search (NAS), which systematically optimized the placement and configuration of these blocks to maximize performance on benchmark datasets.

In summary, the stem layer initiates the feature extraction process with an efficient and compact representation of the input, while the MBConv blocks drive the deeper learning and abstraction of complex patterns. Together, they form the architectural backbone of EfficientNet, enabling it to achieve state-of-the-art results in image

classification tasks with remarkable efficiency, making it highly suitable for deployment in resource-constrained environments.

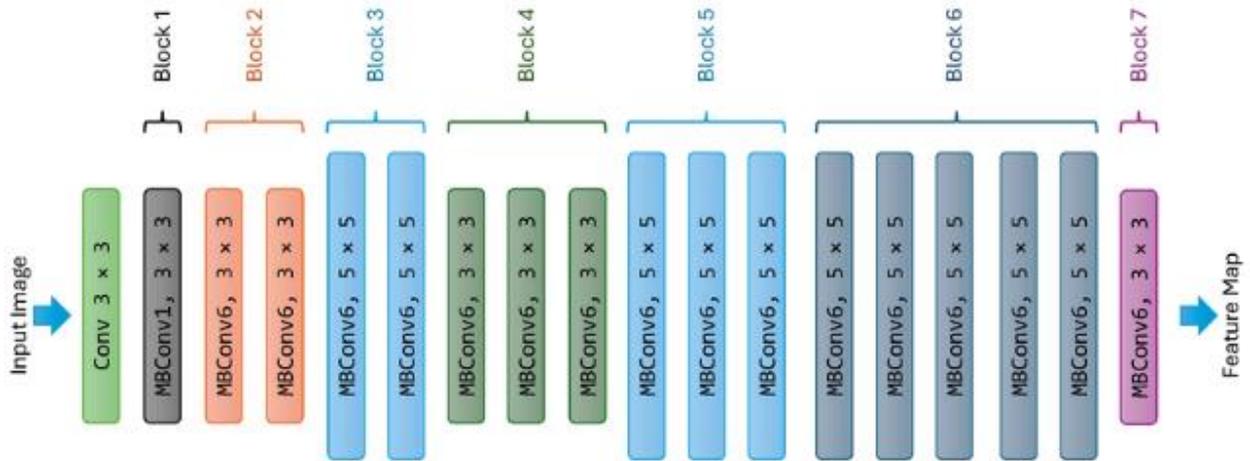
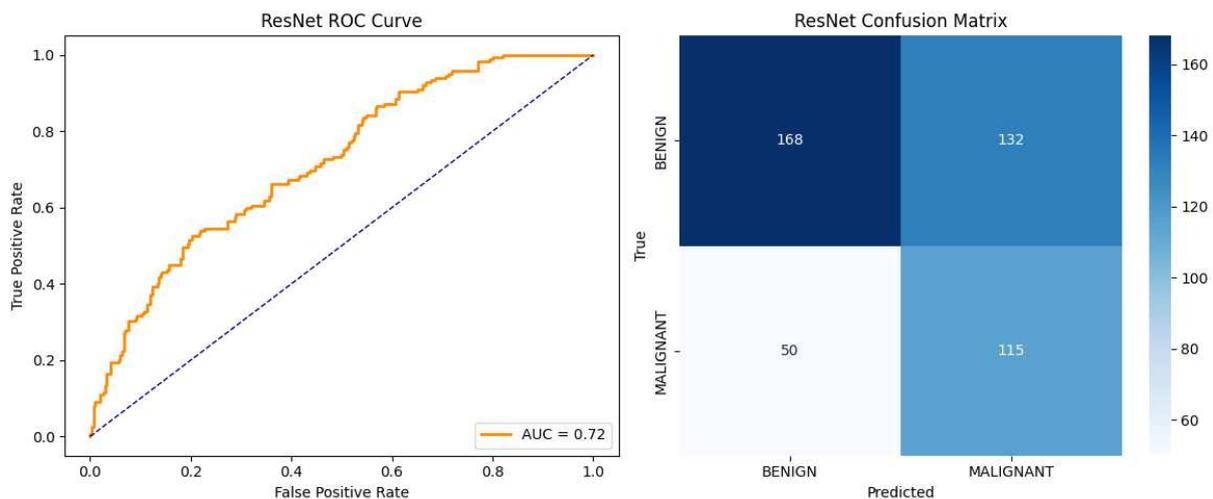
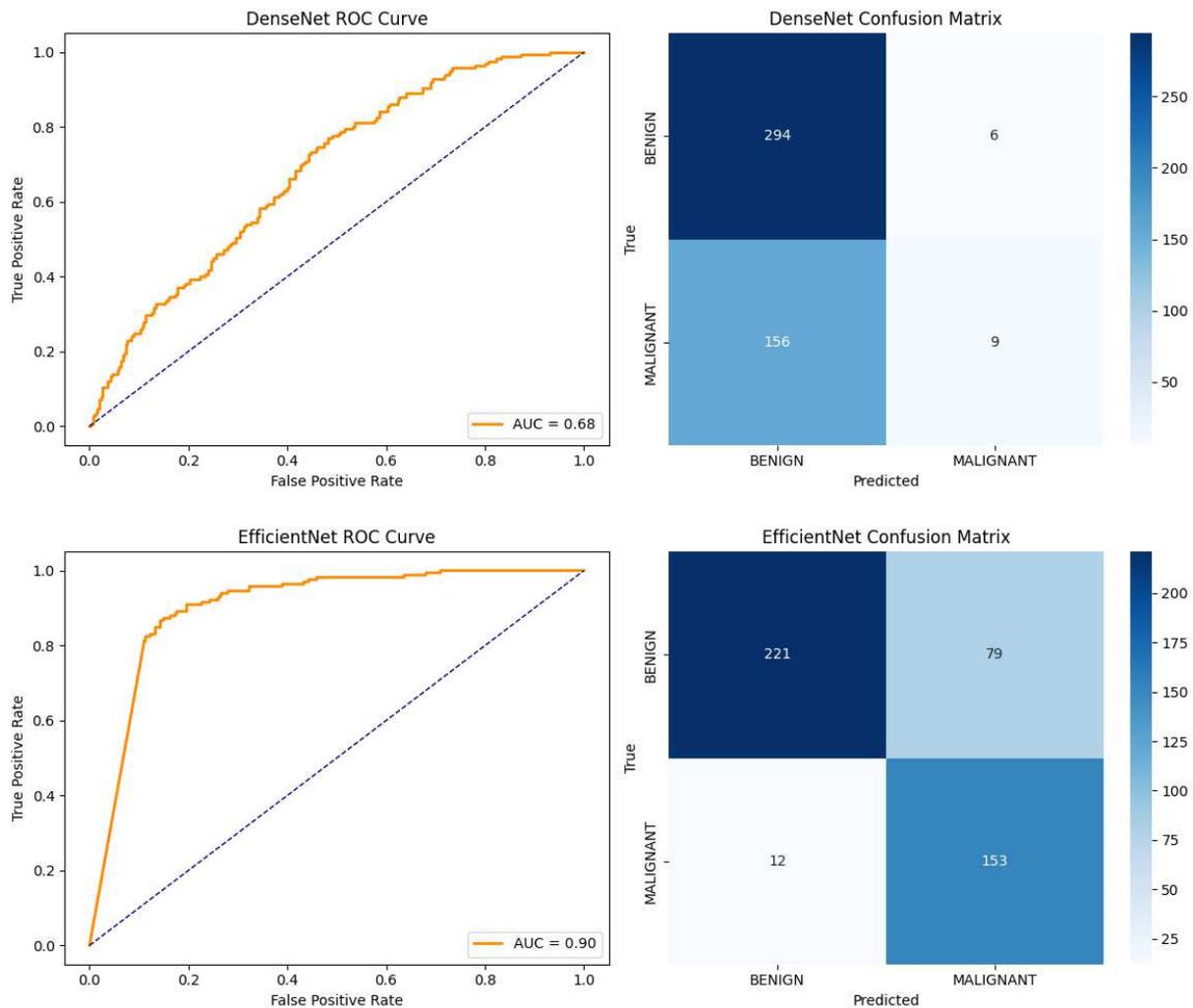


Figure 4.22: EfficientNet-b0 architecture

#### 4.4.4 Experimental evaluation of three Individual CNN models for microcalcifications

Before evaluating the performance of our proposed ensemble models, it is essential to first assess each individual model independently using the CBIS-DDSM dataset. This step ensures a clear understanding of how each backbone, such as ResNet, DenseNet, and EfficientNet, performs on its own in the context of automatic diagnosis of microcalcifications. Testing the models separately allows us to identify their respective strengths and weaknesses, provides a performance baseline, and justifies their inclusion in the ensemble. By establishing the individual capabilities of each architecture, we can better interpret the contribution of each model to the final ensemble and validate whether the ensemble strategy offers meaningful performance improvements over single-model baselines.





**Figure 4.23:** Confusion matrix and ROC curve of ResNet18, DenseNet-121, and EfficientNet-b0

In medical applications, the selection of evaluation metrics is critical, as it directly impacts diagnostic reliability and clinical decision-making. The trade-off between false negatives (missed diagnoses) and false positives (unnecessary follow-ups or interventions) must be carefully balanced to optimize patient outcomes. Sensitivity (Recall) is particularly vital in cases where missing a positive diagnosis could lead to severe consequences, such as delayed cancer detection. Conversely, specificity ensures that healthy patients are not incorrectly classified as having a disease, thereby reducing unnecessary anxiety, additional testing, and overtreatment.

Model	Accuracy	Specificity	Sensitivity	AUC
<b>ResNet-50</b>	60.9%	56.0%	69.7%	0.72
<b>DenseNet-121</b>	65.2%	98.0%	5.5%	0.68
<b>EfficientNet-B0</b>	<b>80.8%</b>	<b>65.9%</b>	<b>94.6%</b>	<b>0.90</b>

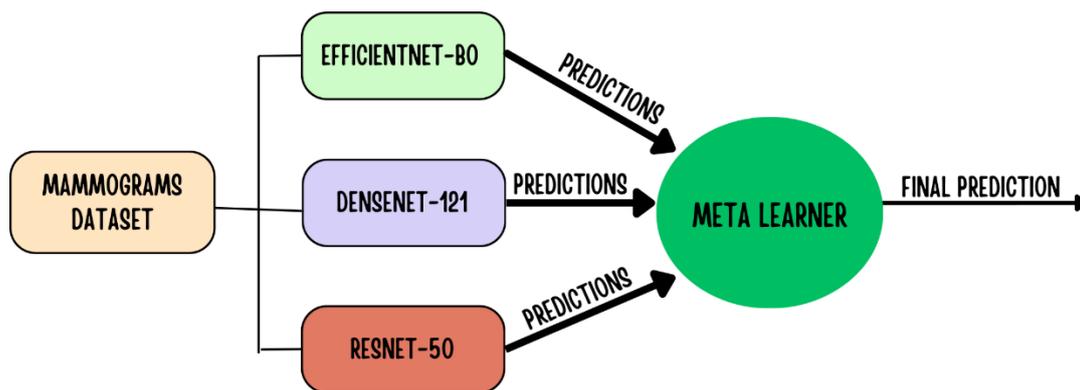
**Table 4.2:** Performance of base-models

The results of the experiments presented in Table 4.2 and illustrated with confusion matrix and ROC curve in Figure 4.23, reveal significant performance differences among the three deep learning models, ResNet-50, DenseNet-121, and EfficientNet-B0, when trained without any augmentation or oversampling on the CBIS-DDSM dataset. ResNet-50 achieved a moderate overall accuracy of 60.9%, with a balanced sensitivity of 69.7% and specificity of 56%. This indicates that ResNet-50 is relatively incapable of detecting benign and malignant cases. On the other hand, DenseNet-121 showed high specificity (98.0%) but an extremely low sensitivity of only 5.45%, which means it accurately identified most benign cases but failed to detect malignant ones. Such behavior suggests a strong bias toward the majority class, likely caused by class imbalance or insufficient feature learning. In contrast, EfficientNet-B0 outperformed both models by a wide margin, achieving an impressive accuracy of 80.43%, with high sensitivity (92.73%) and low specificity (65.9%). It demonstrated strong performance in identifying malignant lesions but with high number false negative cases.

The results obtained from these experiments will serve as a baseline for evaluating the performance of the two proposed ensemble models, which are detailed in Sections 4.5 and 4.6. By comparing the individual CNN models with the ensemble approaches, we aim to assess the extent to which model integration improves diagnostic accuracy, robustness, and generalization in the automatic diagnosis of microcalcifications.

## 4.5 Proposed stacked generalization strategy

Stacking, or stacked generalization, involves training several distinct base models on the same dataset. The predictions generated by these base models are then passed to a meta-learner or meta-model [155]. The meta-learner is essential in determining the optimal way to combine the outputs of the base models, analyzing the patterns in their predictions and adjusting the weights accordingly. By learning from the strengths and weaknesses of each individual base model, the meta-learner is able to produce a more accurate final prediction.



**Figure 4.24:** Composition of the proposed stacking system, composed of EfficientNet-b0, DenseNet-121 and ResNet-50 models and optimized meta learner.

In this section we introduce an ensemble learning system using stacking technique. ResNet-50, DenseNet-121, and EfficientNet-B0 models, are trained on the dataset to generate predictions. Subsequently, a meta-learner is proposed to integrate the outputs of each base model, resulting in a final prediction. Figure 4.24 illustrates the composition of the proposed model.

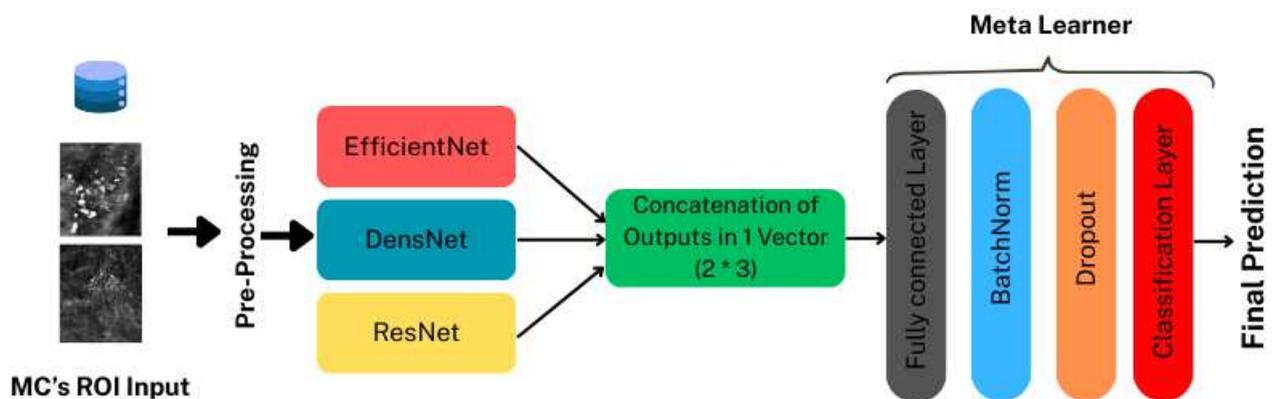
### 4.5.1 Description of stacking model architecture

The proposed ensemble architecture integrates three state-of-the-art pre-trained convolutional neural networks ResNet-50, DenseNet-121, and EfficientNet-B0, to capitalize on their complementary strengths and enhance the classification of microcalcifications. Each of these models is individually fine-tuned for the specific task by adapting their classification heads to output predictions across two target classes: benign and malignant. Rather than relying on the decision of a single model, this ensemble strategy seeks to aggregate multiple perspectives to improve performance.

Following individual inference, the output logits from each model, each consisting of a two-dimensional vector representing the class scores are concatenated to form a unified feature vector of size 6 (i.e., 2 outputs  $\times$  3 models). This merged feature representation serves as the input to a meta-learner, which is responsible for synthesizing the predictions and producing a final, refined classification.

The meta-learner functions as a high-capacity decision-making layer designed to model the complex interactions and dependencies among the outputs of the base models. Architecturally as illustrated in Figure 4.25, it is implemented as a multilayer perceptron. The first layer is a fully connected (Linear) layer with 256 hidden units, which projects the 6-dimensional input to a higher-dimensional feature space. This is followed by a ReLU activation function to introduce non-linearity, enabling the model to learn complex relationships between features.

To further stabilize and accelerate the training process, Batch Normalization (BatchNorm1d) is applied after the activation, normalizing the activations and reducing internal covariate shift. To mitigate overfitting, a Dropout layer with a probability of 0.5 is employed, randomly deactivating neurons during training and encouraging the network to learn more robust representations.



**Figure 4.25:** Architecture of the proposed stacking model with Meta-learner.

Subsequently, a second Linear layer with 128 hidden units is introduced, again followed by ReLU activation and Batch Normalization. This intermediate layer allows the model to refine its understanding of the latent patterns embedded within the combined

predictions. Finally, the meta-learner concludes with a Linear output layer that maps the 128-dimensional feature space to a 2-dimensional output corresponding to the two classes. A Softmax activation function is applied to convert the raw scores into normalized class probabilities, enabling probabilistic interpretation of the final prediction.

### 4.5.2 Experiments

The experiments in this study were conducted within the Kaggle environment, a cloud-based platform that offers powerful GPU support and an accessible interface for data science and machine learning tasks. Kaggle provides pre-configured environments with popular libraries, making it an ideal choice for deep learning research and experimentation. The implementation of deep learning models was carried out using PyTorch, a widely used open-source deep learning framework known for its dynamic computation graph, ease of debugging, and strong community support. In addition to PyTorch, several auxiliary libraries were utilized to streamline the development process, including Torchvision for handling image transformations and pretrained models, Pandas for data manipulation, NumPy for numerical operations, Matplotlib and Seaborn for data visualization, and Scikit-learn for evaluation metrics and data splitting. This robust combination of tools facilitated efficient model training, evaluation, and experimentation in a reproducible and scalable manner.

Given the importance of early and accurate diagnosis, accuracy alone is insufficient as a standalone metric in medical applications. A high accuracy may still conceal imbalances in sensitivity or specificity, which could lead to clinically significant errors. For instance, a model with high accuracy but low sensitivity may fail to detect a significant number of malignant cases, compromising patient safety. Similarly, a model with low specificity may trigger an excess of false positives, increasing the burden on healthcare systems and causing patient distress. Thus, sensitivity and specificity must always be evaluated alongside accuracy to ensure that the model performs reliably in a real-world clinical setting.

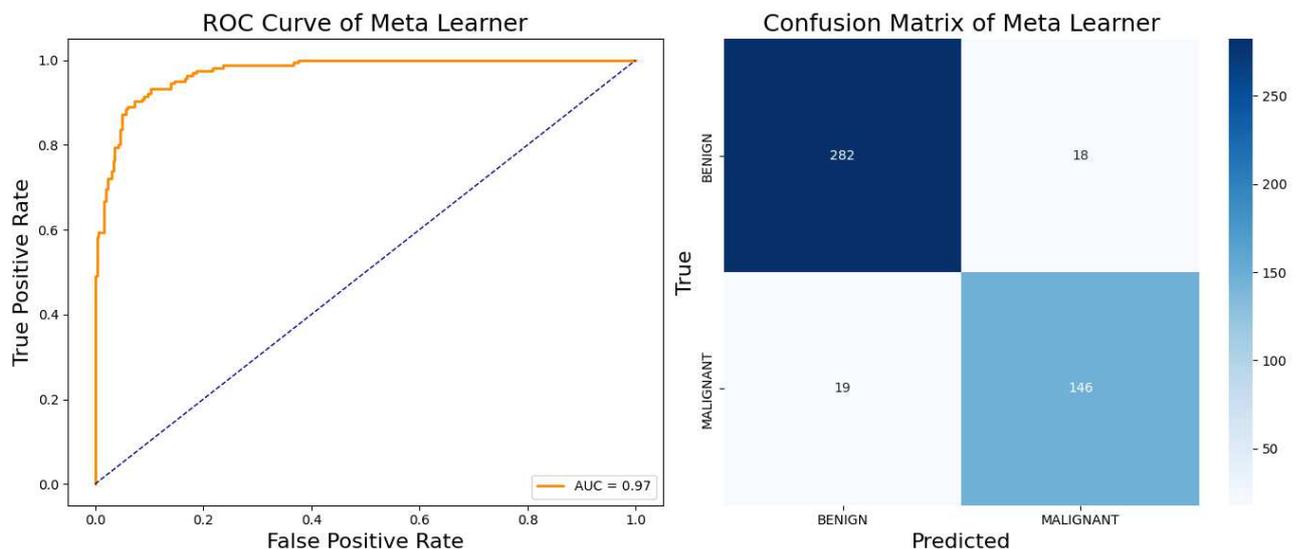
In this experiments, the model's performance is evaluated using accuracy, sensitivity, specificity, precision, and the F1-score to provide a comprehensive assessment of its classification ability. The confusion matrix (Figure 4.26) highlights the model's effectiveness in distinguishing malignant from benign cases, achieving a high overall accuracy of 92.04%. The model correctly classifies 146 malignant cases (true positives) and 282 benign cases (true negatives), while 18 benign cases were misclassified as malignant (false positives) and 19 malignant cases were misclassified as benign (false negatives). The precision for malignant cases is 94%, meaning that among all cases classified as malignant, the majority are indeed malignant, minimizing the risk of unnecessary interventions. Unexpectedly, the model achieves a low recall (sensitivity) of 88.48% (False negative) that means 18 malignant images were classified as benign, necessitating the introduction of attention mechanisms or weighted samples to make models focus more on malignant samples. The specificity of 94% further confirms that the meta-learner effectively identifies benign cases, minimizing false positives. The F1-score

of 95.48% reinforces the model's ability to maintain a balance between precision and recall, ensuring reliable classification across both classes.

Model	Accuracy	Specificity	Sensitivity	AUC
<b>Our Proposed Stacking system</b>	<b>92.04%</b>	<b>94.00%</b>	<b>88.48%</b>	<b>0.97</b>
<b>ResNet-50</b>	60.90%	56.00%	69.70%	0.72
<b>DenseNet-121</b>	65.20%	98.00%	5.50%	0.68
<b>EfficientNet-B0</b>	<b>80.80%</b>	<b>65.9%</b>	<b>94.6%</b>	<b>0.90</b>

**Table 4.3:** Comparison between the proposed stacking system and individual base-models performances.

The results presented in the Table 4.3 demonstrate the superior performance of the proposed stacking ensemble model compared to the individual CNN models. The stacking system achieved the highest accuracy (92.04%) and AUC (0.97), indicating strong overall classification performance and excellent discriminative capability. While EfficientNet-B0 showed relatively high sensitivity (94.6%), its lower specificity (65.9%) suggests a tendency to produce more false positives. Conversely, DenseNet-121 achieved very high specificity (98.0%) but extremely low sensitivity (5.5%), making it unsuitable for detecting true positive cases. ResNet-50 showed more balanced but modest performance across all metrics. These results highlight that the ensemble model effectively combines the strengths of individual networks, resulting in better performance in diagnosing microcalcifications



**Figure 4.26:** Confusion matrix and ROC Curve of the Stacking Ensemble model proposed with optimal Meta Learner.

Further performance validation is provided by the Receiver Operating Characteristic (ROC) curve (Figure 4.26), which plots the true positive rate (sensitivity) against the false positive rate ( $1 - \text{specificity}$ ) at different classification thresholds. The model achieves an

Area Under the Curve (AUC) of 0.97, indicating exceptional discriminatory ability between benign and malignant cases. An AUC close to 1 signifies that the model is capable of achieving high sensitivity while maintaining minimal false positives, a key requirement in medical diagnostics. The steep initial rise of the ROC curve further confirms that the model rapidly captures positive cases with minimal error, reinforcing its clinical viability for automated microcalcification detection. Overall, these results highlight the robustness and reliability of the proposed meta learner, making it a valuable decision-support tool in radiology and breast cancer screening.

Authors	Dataset	Models	Acc	Spec	Sens	Auc
<b>Our Proposed Stacking Model</b>	<b>CBIS-DDSM</b>	<b>ResNet-50, EfficientNet-B0, and DenseNet-121</b>	<b>92,04</b>	<b>94,00</b>	<b>88,48</b>	<b>0,97</b>
Kang et al. [157]	DDSM	ResNet-101, Xception, Inception-v3, InceptionResNet-v2, and DenseNet-201, alongside an ensemble model for classification	81,54	91,41	82,47	0,85
Gerbasi et al. [158]	INbreast	DeepMica	83	-	-	0,89
Singh et al. [163]	CBIS-DDSM	InceptionResNetV2, experimented with four optimizers (ADAM, ADAGrad, ADADelta, RMSProp)	94	80	97	0.96
Cai et al. [155]	INbreast	DCNN	83.7	-	90	0,94
Teoh et al. [160]	DDSM-MIAS	AlexNet, GoogLeNet, VGG16, and ResNet-50	93.05	-	-	-
Kumar et al. [171]	CBIS-DDSM	Inception and ResNetV2	94		97	96
Rehman et al. [172]	CBIS-DDSM	FC-DSCNN with the DCNN	90	82	99	-
Chen et al. [173]	DDSM	Mammogram classification using fine-tuning of ResNet	-	-	93.83	-
Young et al. [174]	Private	EfficientNet and DenseNet	95	87	88	-

**Table 4.4:** Comparison between proposed model and some previous works.

As presented in Table 4.4, the superior performance of the proposed model is largely attributed to the adoption of an ensemble learning strategy, which integrates the strengths of three state-of-the-art convolutional neural network architectures: ResNet-50,

EfficientNet-B0, and DenseNet-121, in conjunction with a meta-learner. This architectural design allows the model to exploit complementary features learned by the individual base models, leading to enhanced feature representation, improved generalization, and increased classification accuracy. The resulting ensemble model demonstrates exceptional performance, achieving an accuracy of 92.04% with 94% of specificity, an Area Under the Curve (AUC) of 0.97, and a sensitivity of 88.48%, clearly outperforming more conventional approaches such as those proposed by Kang et al. and Gerbasi et al. in the domain of microcalcification classification.

An innovative aspect of this ensemble framework is the use of a Fully connected layers within the meta-learner, which, although less common in scenarios where the inputs are soft predictions rather than spatial data, offers notable advantages. Traditional Fully connected layers are designed to capture spatial patterns in data such as images; however, in this context, meta learner operations are leveraged to detect local dependencies and correlations among the predictions of the base models. These base models may exhibit similar error patterns under specific conditions, suggesting that their outputs are not entirely independent. By applying multilayer perceptron over the concatenated predictions, the meta-learner is able to capture structured relationships and contextual patterns among the outputs, thereby improving its ability to assign adaptive weights to each model's contribution in various input scenarios. This enables the meta-learner to dynamically calibrate its decision-making process, prioritizing the most reliable sources of information depending on the input characteristics. As a result, the integration of MLP block enhances the expressive capacity and discriminatory power of the meta-learner, contributing to the overall robustness and effectiveness of the proposed diagnostic framework.

## 4.6 Proposed boosting strategy

To address the problem of false negative and false positive rates in classification of microcalcifications, boosting strategy is used and evaluated. As explained before in this chapter (see 3.2) Boosting is an ensemble learning strategy designed to transform a collection of weak learners into a single, strong predictive model by training them sequentially [178]. When applied to ResNet, DenseNet and EfficientNet, a boosting learner constructs multiple CNN models where each successive network is trained to correct the errors made by its predecessors. In our context we will focus more on malignant cases to increase sensitivity. Instead of treating all models equally, the boosting process accentuates instances that were misclassified in previous iterations, allowing subsequent base-models to focus more effectively on difficult cases. By progressively refining the learning process and combining the outputs of all models, the ensemble achieves improved generalization and enhanced predictive accuracy. This approach leverages the representational power of CNNs while systematically reducing bias and variance through focused, iterative learning.

### 4.6.1 Boosting algorithms in the literature

The training method varies depending on the type of boosting process, known as the boosting algorithm. However, an algorithm generally follows these steps to train the boosting model:

**[1] Step 1**

The boosting algorithm assigns an equal weight to each data sample. It feeds the data into the first machine learning model, known as the base learner. The base learner makes predictions for each data sample.

**[2] Step 2**

The boosting algorithm evaluates the model's predictions and increases the weight of samples with higher prediction errors. It also assigns a weight based on the model's performance. A model that produces highly accurate predictions will have a greater influence on the final decision.

**[3] Step 3**

The algorithm passes the reweighted data to the next model (or base learner).

**[4] Step 4**

The algorithm repeats steps 2 and 3 until the training error falls below a certain threshold.

Different algorithms were developed for boosting ensemble learning, each with its own strategy for improving model performance by focusing on difficult samples. Classical methods such as AdaBoost adjust sample weights based on misclassifications, highlighting hard-to-classify examples in subsequent iterations. Gradient Boosting takes a different approach by training models sequentially to correct the residual errors of the previous models. Other advanced techniques like XGBoost, LightGBM, and CatBoost introduce additional improvements such as regularization, tree pruning, and optimized handling of categorical variables [182, 183]. These algorithms demonstrate the flexibility and power of boosting frameworks to adapt to a wide range of tasks. In the next section AdaBoost, Gradient Boost and XGBoost will be detailed

**4.6.1.1 Adaptive Boosting (AdaBoost)**

Adaptive Boosting (AdaBoost) is one of the earliest boosting models developed. It adapts and attempts to self-correct at each iteration of the boosting process [179].

Initially, AdaBoost assigns equal weights to all data samples. After each base learner, it automatically adjusts the weights of the data points. Samples that are misclassified are given more weight to correct them in the next round. This process is repeated until the residual error, the difference between the true and predicted values falls below an acceptable threshold.

AdaBoost can be used with many types of predictors and is generally less sensitive to noise compared to some other boosting algorithms. However, it may perform poorly when there is high feature correlation or high data dimensionality. Overall, AdaBoost is well-suited for classification problems.

**4.6.1.2 Gradient Boosting (GB)**

Gradient Boosting (GB), also called Gradient Boosted Machines, is similar to AdaBoost in that it also relies on a sequential training technique. However, the key difference is that GB does not explicitly increase the weight of misclassified points. Instead, GB optimizes a

differentiable loss function by training new base learners sequentially, each one attempting to correct the errors (residuals) of its predecessor [180].

Rather than adjusting weights, GB fits each new model to the residual errors of the combined previous models. This strategy aims to generate accurate results from the beginning, instead of correcting errors as in AdaBoost. As a result, GB can produce highly accurate models. Gradient Boosting can be used for both classification and regression problems.

#### 4.6.1.3 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) improves Gradient Boosting in terms of computational speed and scalability. XGBoost utilizes multiple CPU cores to enable parallel learning during training. It is capable of handling very large datasets, making it particularly attractive for big data applications [181].

XGBoost increases performance through several techniques, including parallelization, distributed computing, cache optimization, and out-of-core processing (handling data too large to fit into memory). These improvements make XGBoost one of the fastest and most scalable boosting algorithms available.

#### 4.6.2 Proposed CSB-EWA boosting algorithm

In this section a new boosting algorithm named Cost-Sensitive Boosting with Error Weighted Adjustments (CSB-EWA) is proposed to address the problem of false positive and false negative rate when base-models were tested in section 4.4.4 of this chapter. In traditional AdaBoost, the weight update for each sample is based on whether the sample is correctly classified or misclassified. However, in this proposition, the algorithm adapts this weight update to prioritize different types of errors based on the False Positive Rate (FPR) and False Negative Rate (FNR). This is especially useful in contexts like medical images classification, where the cost of false negatives (missed diseases) and false positives (wrongly diagnosing a disease) are not equal. By this algorithm we aim to create an ensemble learning capable to well predict malignant microcalcifications. In fact, the three models used in this study act differently with CBIS-DDSM dataset as presented in Table 4.2, DenseNet-121 has very high specificity (98.00%) but very low sensitivity with 5.45%, in the other hand EfficientNet-b0 predict malignant cases with high rate (Sensitivity of 92.73%) but struggle with benign images. To overcome this problem the proposed algorithm, adjust samples weight's according to false positive rate and false negative rate of each model to increase the performances. The CSB-EWA boosting algorithm illustrated in Figure 4.27 is processed as follows:

- 1- Initial Weight Assignment:** Initially, all samples are given equal weights, typically set to  $w_i = \frac{1}{N}$ , where  $N$  is the total number of samples. These weights are updated after each iteration to focus more on difficult or misclassified samples.
- 2- Training the Base Learner:** Base-models or classifiers are trained on the weighted dataset, and predictions  $h_t(x)$  are made for each sample. The base learner's accuracy, sensitivity and specificity are evaluated, and the weighted error rate is calculated.

These rates measure how well the model is handling misclassifications, focusing on positive and negative errors, respectively.

### 3- Calculate the Misclassification Error (Base Learner Error):

The base learner error  $\epsilon_t$  is the weighted error of the classifier for each iteration  $t$ , where the weights of misclassified samples are considered. It is computed as:

$$\epsilon_t = \frac{\sum_{i=1}^n w_i * I(y_i \neq \hat{y}_i)}{\sum_{i=1}^n w_i}$$

where:

- $w_i$  is the weight of sample  $i$ .
- $y_i$  is the true label of sample  $i$ .
- $\hat{y}_i$  is the predicted label of sample  $i$ .
- $I(\cdot)$  is an indicator function that is 1 if the condition is true (i.e., misclassification) and 0 otherwise.

### 4- Adjust Penalties ( $\lambda$ ) Based on FPR and FNR: The penalties $\lambda_{FP}$ and $\lambda_{FN}$ are adjusted based on the FPR and FNR to focus more on the types of errors that are higher in the classifier's performance:

- $\lambda_{FN} = 1 + \gamma \cdot \text{FNR}$
- $\lambda_{FP} = 1 + \gamma \cdot \text{FPR}$
- Here,  $\gamma$  (gamma) is a scaling factor that controls how much influence FPR and FNR have on the penalty adjustment (e.g.,  $\gamma=2$  or 3 or 5).

### 5- Adjust the Weight Update: Instead of updating weights uniformly based on misclassification alone as done by AdaBoost, the weight update is adjusted based on the FPR and FNR. If the FNR is high (i.e., the model is missing a significant number of positive cases), the weights for false negatives are increased more significantly to force the model to focus more on correcting these errors in the next iteration. Similarly, if the FPR is high (i.e., the model is wrongly classifying too many negatives as positives), the weights for false positives are adjusted accordingly. This dynamic adjustment ensures that the algorithm prioritizes the errors that matter more for the problem of microcalcifications, such as reducing false negatives in medical image classification. The weight error $w_i(t+1)$ for sample $i$ after the $t$ -th base learner is updated using the error, as follows:

**For False Negatives (FN):** If the sample is a false negative (i.e., predicted negative but actual positive), the weight is updated based on the FN penalty:

$$w_i(t+1) = w_i(t)e^{(\lambda_{FN} \cdot \alpha_t)}$$

where  $\alpha_t$  is the weight of the base learner at iteration  $t$ , typically computed using:

$$\alpha_t = \frac{1}{2} \ln\left(\frac{1 - \epsilon_t}{\epsilon_t}\right)$$

**For False Positives (FP):** If the sample is a false positive (i.e., predicted positive but actual negative), the weight is updated based on the FP penalty:

$$w_i(t + 1) = w_i(t)e^{(\lambda_{FP} \cdot \alpha_t)}$$

**For Correctly Classified Samples:** If the sample is correctly classified, its weight is updated with a negative exponential factor:

$$w_i(t + 1) = w_i(t)e^{-\alpha_t}$$

**Normalization:** of all weights so the sum of  $w_i=1$

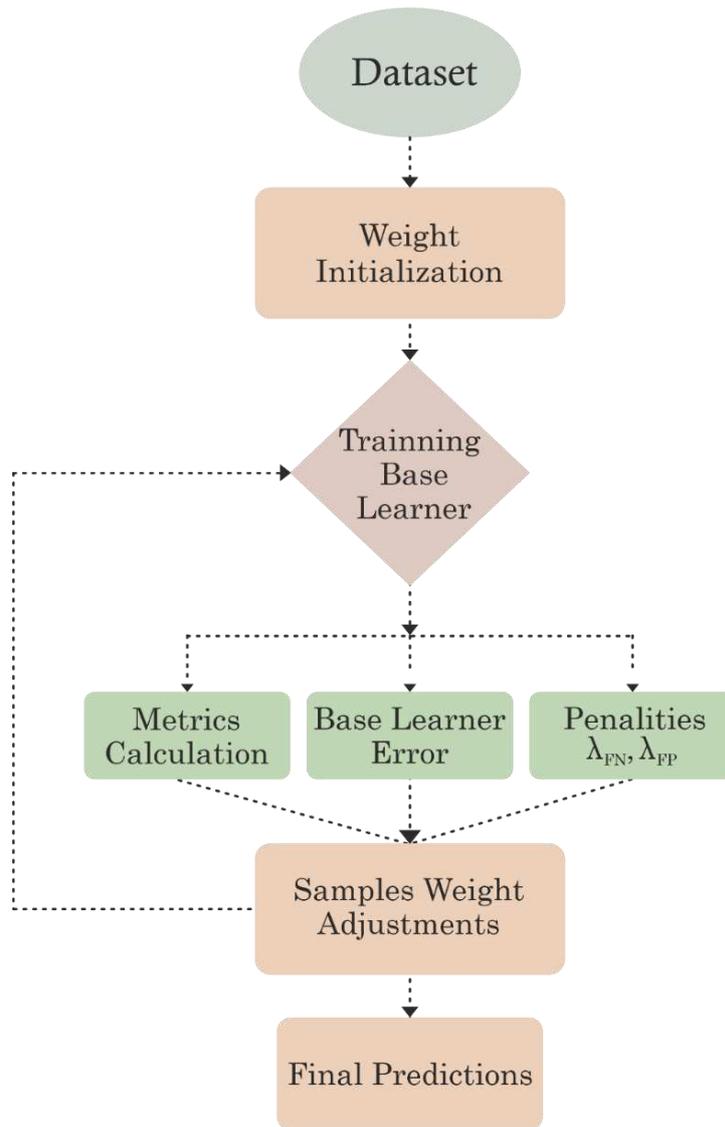
**6- Go to step 2**

**7- Final Prediction in CSB-EWA:** After training all base-models, for each sample the final prediction  $H(x)$  is computed as:

$$H(x) = \text{sign} \left( \sum_{t=1}^T \alpha_t h_t(x) \right)$$

Where :

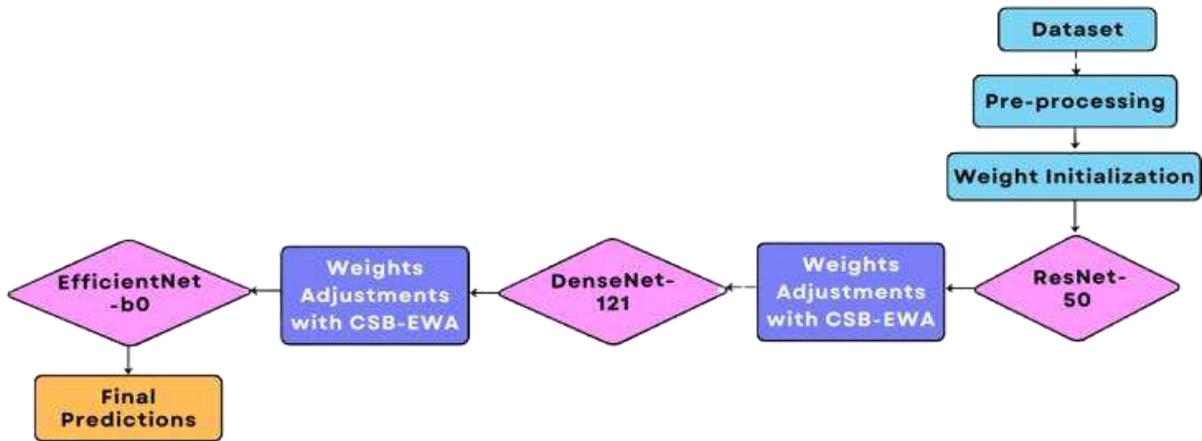
- $h_t$  is the prediction of the t-th base model (usually outputs +1 or -1),
- $\alpha_t$  is the importance weight of the t-th model (depends on FPR, FNR, error, etc.),
- The sign function returns +1 if the weighted sum is positive, -1 otherwise.



**Figure 4.27:** Algorithm diagram of Cost-Sensitive Boosting with Error Weighted Adjustments.

CSB-EWA algorithm is integrated into ensemble learning system to improve the classification of microcalcifications. Initially, all training samples are assigned equal weights. The first base learner, ResNet-50, is trained using these weights, and its performance is evaluated to calculate a weighted misclassification error or  $\epsilon_t$ . Based on the classifier's FPR and FNR, penalty terms ( $\lambda_{FP}$  and  $\lambda_{FN}$ ) are calculated using a scaling factor  $\gamma$  to emphasize the correction of more critical errors, particularly false negatives. Weights of misclassified samples are then updated accordingly: false negatives and false positives receive exponentially increased weights proportional to their penalties, while correctly classified samples are penalized less. The updated weights are normalized and passed to the second base learner, DenseNet-121, and subsequently to the third model EfficientNet-B0. After the three models are trained, final predictions are made using a weighted majority vote, where each model's prediction is scaled by its individual performance ( $\alpha_i$ ). This adaptive strategy ensures that the ensemble focuses on minimizing clinically significant misclassifications. As illustrated in Figure 4.28 CSB-EWA algorithm

orchestrates the entire system of ensemble learning, from the weights initialization to final predictions.



**Figure 4.28:** Boosting ensemble learning system using CSB-EWA algorithm.

### 4.6.3 Experiments and results

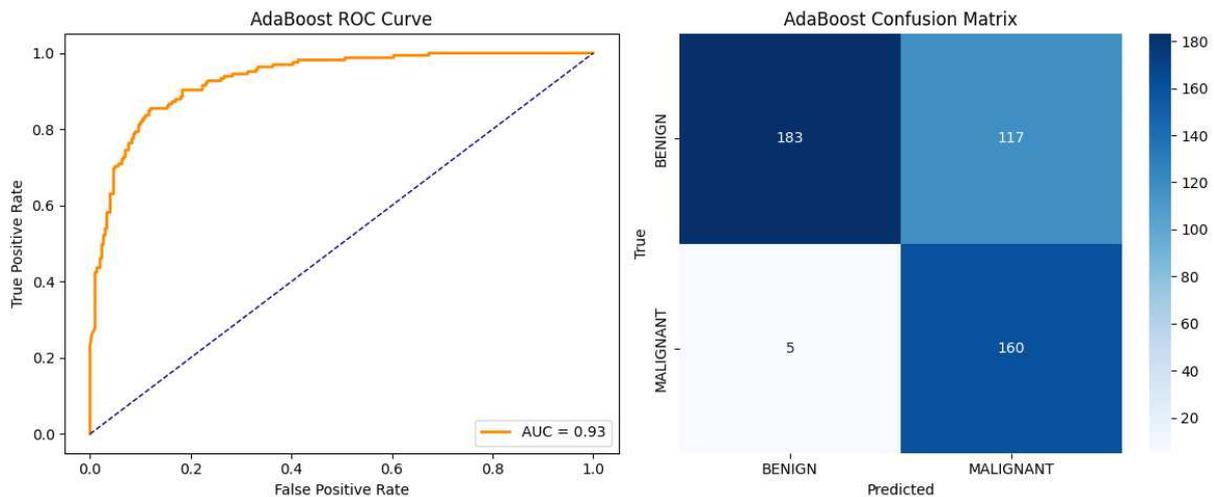
To evaluate the performance of the proposed CSB-EWA (Cost Sensitive Boosting with Error Weighted Adjustments) algorithm, we conducted extensive experiments on the CBIS-DDSM dataset. The dataset contains annotated mammography images categorized into benign and malignant classes. The input images are resized to  $224 \times 224$  pixels to maintain a balance between computational efficiency and model performance. A size of  $224 \times 224$  is a common choice in deep learning, especially for pre-trained models like ResNet, DenseNet, and EfficientNet, which are designed to accept this input dimension as most of the architectures were trained on datasets like ImageNet, where images are also  $224 \times 224$ . This resolution is large enough to preserve significant features like tiny calcium deposits of calcifications. Given the training setup with the Adam optimizer, learning rate of  $1 \times 10^{-4}$ , cross-entropy loss, a batch size of 16, and 30 epochs on an NVIDIA Tesla P100 GPU (Kaggle Environment). We adopted a fixed train-test split strategy, dividing the dataset into 70% training and 30% testing using stratified sampling to preserve class distributions. The test set remained completely unseen during training. The Class distribution of dataset are detailed in section 4.2 of this chapter.

Base Model	Number of Layers	Number of Parameters	Feature Vector Size (Output)
ResNet-50	50	~25.6 million	2048 (for a $224 \times 224$ input)
DenseNet-121	121	~8.0 million	1024 (for a $224 \times 224$ input)
EfficientNet-B0	82	~5.3 million	1280 (for a $224 \times 224$ input)

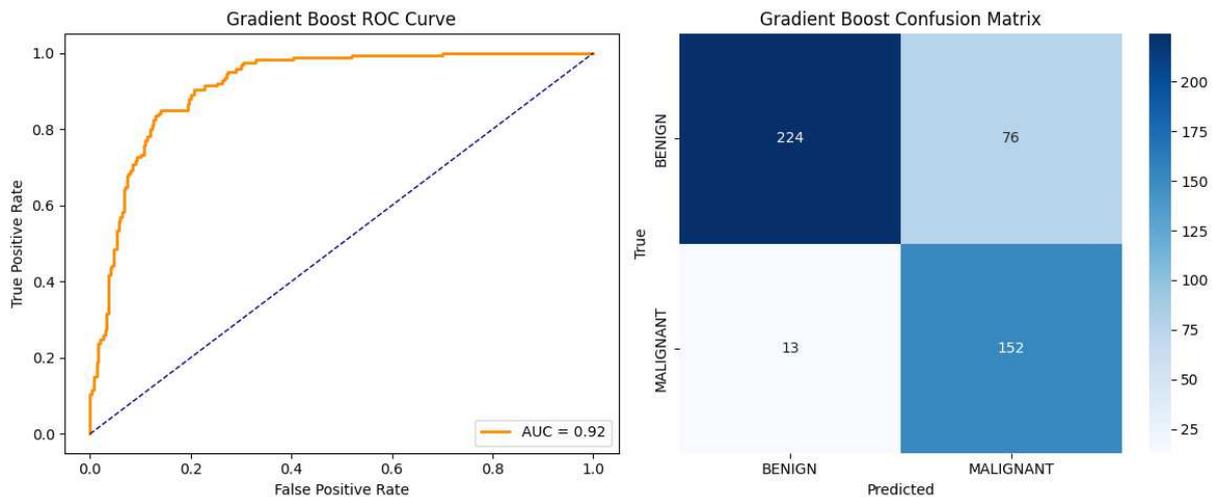
**Table 4.5:** Different characteristics of models used in our experimentation.

The table presents a comparative overview of the three base CNN architectures employed in this study, ResNet-50, DenseNet-121, and EfficientNet-B0, highlighting their structural complexity and output characteristics. ResNet-50, with 50 layers and approximately 25.6 million parameters, is the most computationally intensive model, producing a 2048-dimensional feature vector. DenseNet-121, while deeper in terms of layer count (121 layers), is more parameter-efficient with around 8 million parameters, yielding a 1024-dimensional output. EfficientNet-B0 offers a balanced architecture with around 82 layers and only 5.3 million parameters, yet produces a high-dimensional 1280-feature output due to its compound scaling strategy. These differences illustrate the trade-offs between model depth, parameter efficiency, and representational capacity, which are leveraged in the ensemble to capture diverse and complementary feature representations.

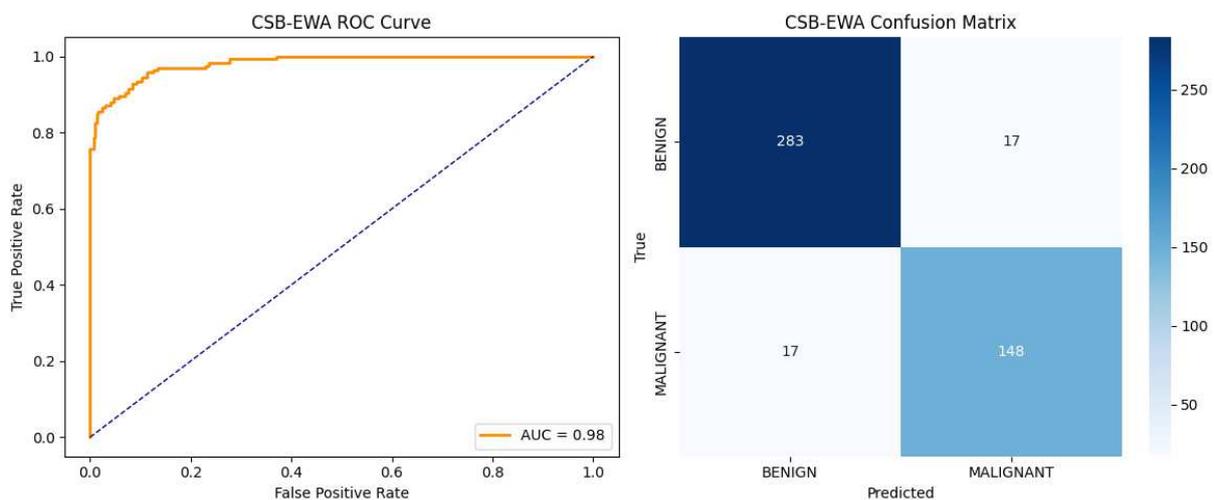
In our boosting models, learning is carried out iteratively, where each new model is trained to correct the mistakes of the previous one. Initially, all samples are weighted equally, and a base model is trained on the dataset. After each iteration, misclassified samples receive higher weights, making them more influential in the next round. In the CSB-EWA strategy, base learners, ResNet101, DenseNet-121, and EfficientNetB7, are trained sequentially with the parameters summarized in Table 4.4. The weighted error rate is computed for each base learner, and penalties for false positives (FPR) and false negatives (FNR) are dynamically adjusted using scaling factors ( $\lambda_{FP}$  and  $\lambda_{FN}$ ). These penalties are incorporated into a refined weight update formula, ensuring the model focuses more on critical misclassifications, such as false negatives, which are particularly important in medical image analysis like microcalcifications. After training, the final ensemble prediction is obtained through a weighted sum of the individual models' outputs, with the sign function used to assign the final class label.



**Figure 4.29:** Results obtained with first scenario: Boosting of the ensemble learning using Adaboost.



**Figure 4.30:** Results obtained with second scenario: Boosting of the ensemble learning using Gradient Boost.



**Figure 4.31:** Results obtained with third scenario: Boosting of the ensemble learning using proposed CSB-EWA algorithm.

To evaluate the effectiveness of the CSB-EWA approach, three scenarios was tested and compared using classical boosting techniques, AdaBoost and Gradient Boosting. The results are summarized in Table 4.5, with confusion matrices and ROC curves presented in Figures 4.29, 4.30, and 4.31.

The first scenario used Adaboost algorithm to test the ensemble learning using ResNet-50, DenseNet-121 and EfficentNet-b0. In this experiments sensitivity is improved to 96.97% but compromised specificity at 61%. Adaboost algorithm could not decrease false positive and false negative rates in same time. The main reason is that Adaboost adjust all misclassified samples with same weights rather than false negative or false positive samples. In imbalanced dataset like CBIS-DDSM these adjustments push models to focus more on dominant class, whatever they are misclassified as false negative or false positive, do the necessity to make relative adjustments as proposed in CSB-EWA algorithm.

The second scenario in these experiments use Gradient Boosting in the ensemble system. Contrary to Adaboost Gradient Boost improved sensitivity to 92,12% and perform

better than Adaboost with 80.75% accuracy and a 0.92 AUC. But specificity still low with 74,67%. In a binary classification task like our problematic, each sample is associated with a predicted probability representing the likelihood of belonging to the positive class. Gradient Boost evaluates how far each prediction deviates from the actual class label (0 or 1), and these differences are used to compute the residuals or pseudo-residuals. These residuals reflect the errors of the current model and guide the training of the next learner, which is specifically designed to correct these mistakes. In this way, the algorithm implicitly prioritizes samples with larger errors, allowing the model to focus on harder cases. With each iteration, the overall prediction is updated, gradually minimizing the loss function and improving the model's accuracy. The loss function is not representative metric in our context where sensitivity and specificity are the most significant. The results obtained with Gradient boost demonstrated our hypothesis that an ensemble learning using boosting strategy in the context of breast cancer must focus on false positive and false negative samples.

In the third scenario, the CSB-EWA algorithm is applied for the ensemble learning system, and significantly outperformed all others, achieving 92.69% accuracy, 94.33% specificity, 89.70% sensitivity, and an AUC of 0.98. This superior performance is largely due to the dynamic adjustment of sample weights based on FPR and FNR, allowing the model to prioritize correction of the most critical errors. In Figure 4.32 clarified that CSB-EWA achieved a much better balance between sensitivity and specificity, making it the most effective and balanced method compared to the traditional boosting approaches.

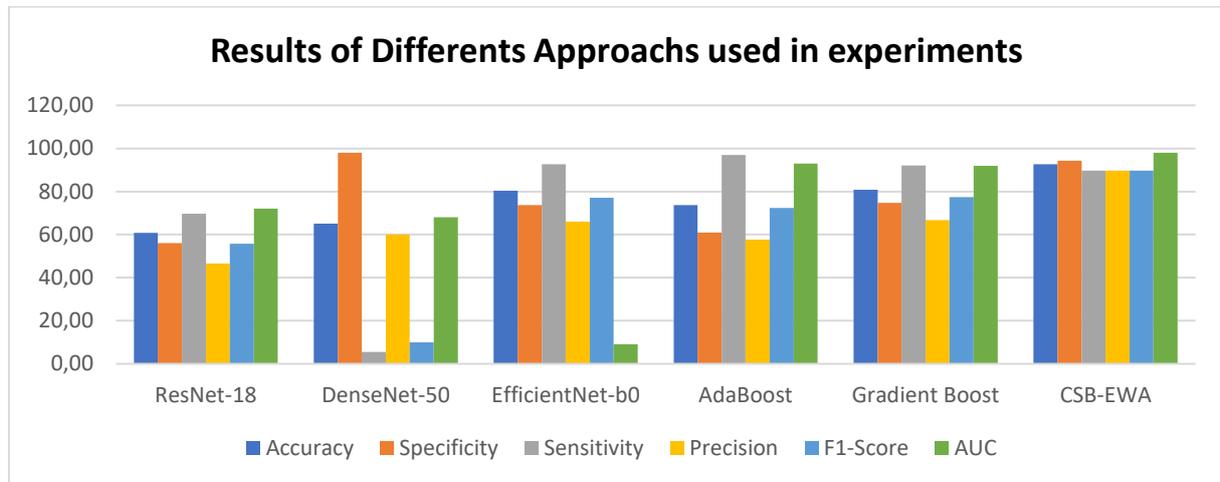
	<b>Accuracy %</b>	<b>Specificity %</b>	<b>Sensitivity %</b>	<b>Precision %</b>	<b>F1-Score %</b>	<b>AUC</b>
ResNet-50	60,86	56,00	69,69	46,52	55,81	0,72
DenseNet-121	65,16	98,00	5,45	60,00	10,00	0,68
EfficientNet-b0	80,43	73,67	92,73	65,95	77,10	0,90
AdaBoost	73,76	61,00	96,97	57,76	72,40	0,93
Gradient Boost	80,75	74,67	92,12	66,67	77,34	0,92
<b>Proposed CSB-EWA</b>	<b>92,69</b>	<b>94,33</b>	<b>89,70</b>	<b>89,70</b>	<b>89,70</b>	<b>0,98</b>

**Table 4.6:** Summary of all metrics and results obtained in experiments process.

The high performance of ensemble learning using CSB-EWA boosting algorithm is due to the separation between false positive and false negative samples after training of each model. Compared to Adaboost that use same weights for all misclassified samples.

More sensitivity is low, meaning high false negative number of samples (High false negative rate). CSB-EWA will attribute high weights to these samples using penalty  $\lambda_{FN}=1+\gamma \cdot FNR$  with  $\gamma$  fixed to 2. High value of  $\gamma$  give more influence to false negative and false positive penalties and increase more weights samples. These weights adjustments make next learner in the ensemble concentrate more on false negative samples to increase sensitivity if sensitivity is low and focus more on false positive samples if specificity is low. This approach makes the system converge to a balanced performance between specificity and

sensitivity. This balancing was impossible to achieve using individual models or boosting using Adaboost or Gradient Boost as shown in Figure 4.32.



**Figure 4.32:** Histogram of metrics obtained with three scenarios and compared with performance of base-models ResNet-50, DenseNet-121 and EfficientNet-b0.

## 5. Conclusion

This chapter commenced with a comprehensive overview of deep learning principles and a review of recent literature focusing on its applications in medical imaging, particularly in the detection and classification of microcalcifications. Building on this foundation, we introduced a robust ensemble model designed to support radiologists in the accurate differentiation of malignant from benign microcalcifications. The proposed ensemble model integrates three high-performing convolutional neural networks; ResNet-50, EfficientNet-B0, and DenseNet-121. In the first proposed ensemble learning the outputs of base-models are combined via a meta-learner specifically designed to refine final predictions. Unlike conventional ensemble strategies, the meta-learner in our approach incorporates convolutional blocks, allowing it to capture inter-model dependencies and learn more discriminative representations. The model was evaluated on the CBIS-DDSM dataset, achieving outstanding results with an accuracy of 92.04%, AUC of 0.98, specificity of 94%, and sensitivity of 88.48%. Boosting is a very effective strategy used in ensemble learning. In context of breast cancer, generalization of common techniques may not perform as expected. To achieve high performance classification of microcalcifications, another contribution was presented in this chapter with new boosting algorithm named Cost-Sensitive Boosting with Error Weighted Adjustments (CSB-EWA). The main idea was to increase sensitivity and specificity of models with weight samples adjustments based on false positive and false negative rate. These results not only demonstrate the model's consistency and robustness in clinical scenarios but also emphasize its potential to reduce false negatives and unnecessary follow-ups. The chapter concludes by highlighting the effectiveness of ensemble learning in medical image classification and suggesting further improvements through extended datasets and architectural refinements. Also, the problem of complexity should be addressed using less complicated CNNs models to reduce time execution and possibility to deployed such solutions in embedded systems like smart phones.

## 6. Conclusion and Perspectives

Our thesis addressed the problem of feature extraction and classification for automatic breast cancer diagnosis. Breast cancer represents a major public health concern due to its high prevalence and significant impact on mortality worldwide. Early detection and diagnosis of breast cancer became key procedures for improving patient health, as they significantly increase the chances of successful treatment and long-term survival. Identifying the disease at an early stage allows for less aggressive therapies, reduces the risk of metastasis, and lowers treatment costs. Breast imaging modalities such as mammography, ultrasound, MRI, and advanced techniques like MBI and PET play a critical role in the early detection and diagnosis of breast cancer. However, interpreting these images can be challenging for radiologists due to the complexity of breast tissue, the subtlety of early signs like masses and microcalcifications, and variations in image quality. These difficulties can lead to missed or inaccurate diagnoses. Integrating CAD systems helps address these challenges by enhancing image interpretation, reducing human error, and supporting more accurate and consistent clinical decisions.

Two main approaches were used in CADx systems since the emergence of artificial intelligence, machine learning and deep learning, each offering distinct strengths and weaknesses. Traditional machine learning approaches rely on handcrafted features, which can be effective for well-defined patterns but often struggle with complex or subtle variations in medical images, often limiting their generalizability across diverse datasets.

Features extraction represent a very important step in CADx systems process. In fact, high quality of features affects the accuracy and efficiency of the entire system. In this thesis, our first contribution addresses the problem of description of spiculated masses which represents a high risk of malignancy. In the literature many descriptors have been designed and performs well in describing normal masses but fails considerably against irregular masses and specially with high spicules. We have remarked that most of techniques and descriptors describe the whole mass, however, radiologists focus on the irregularities and spiculated parts of masses.

Our first contribution in this thesis, consist in proposing a novel descriptor named PATAR (Polygon Approximation Triangle-Area Representation). PATAR focused on separating spiculations formed by concave and convex spaces from the center of masses, then, measure their degree of malignancy. First, to isolate these sharp lines, a geometric transformation (Polygon Approximation) is applied to enhance important irregularities in contour and ignore small variations and simplify the contour form. The Ramer-Douglas-Peucker (RDP) algorithm was employed to perform polygon approximation on masses. The RDP method is based on the perpendicular distance calculated on the endpoints of a curve (contour in this case), using predefined tolerance parameter  $\epsilon$ . The degree of approximation is controlled by  $\epsilon$ , and optimal

value get to a representative approximation and preserves essential shape characteristics. To well estimate and measure spicules isolated with RDP algorithm, Traingle-Area Representation (TAR Signature) is used to quantify the degree of malignancy of masses by measuring the area of spiculations. Then, creating a vector of features to perform classification. Three classifiers are used, SVM, Random Forest and Fuzzy C-Means to classify CBIS-DDM dataset using 1545 images. The high accuracy obtained prove the success of the strategy adopted that consist on isolate, extract and measure spicules to identify malignant masses.

Convolutional Neural Networks (CNNs), as one of the most successful deep learning models, have gained a prominent place in the field of computer vision due to their exceptional ability to automatically extract and learn spatial features from image data. CNNs network combined with Ensemble Learning techniques such as Stacking (Stacked Generalization) can improve significantly the performance of CADx systems. Staking technique use meta-learner (also called a meta-model) to combine the outputs of multiple base models and make the final predictions.

The second contribution in this thesis comprise the conception of an optimal meta-learner using convolutional network to combine the predictions of three models, ResNet-50, DenseNet-121 and EfficientNet-b0. Even the use of convolutional block to learn from predictions issued from base-models seems unusual since CNNs networks deal with spatial data like images, the results obtained in classification of microcalcifications demonstrate the efficiency of the approach. The meta-learner is capable of learning correlation between predictions and detecting patterns to make final decision. In addition of stacking ensemble learning, boosting strategy enchain models sequentially, and adjust samples weights after each iteration based on the error of each model and try to focus more on misclassified images by attributing more weights to these samples.

The third contribution presented in this work, introduced a novel boosting algorithm to adjust weights based on the false positive and false negative rates. The high performance of CSB-EWA is due essentially to  $\lambda_{FP}$  and  $\lambda_{FN}$  penalties used to make models focus more on false positive and false negative samples, instead of assign same weight to all misclassified samples as done with AdaBoost. CSB-EWA algorithm is capable of achieving high performance with balance between specificity and sensitivity. ResNet-50, DenseNet-121 and EfficientNet-b0 was trained sequentially and CSB-EWA adjust samples weights after each model. The obtained results using CBIS-DDSM, proves that using ensemble learning is very benefic in complex tasks like classification of microcalcifications.

As perspectives we envisage as an extension of this thesis, developing a model capable of surpassing the problem of generalization in breast cancer datasets. Because, in feature-based approaches and deep learning models, the features (e.g., shape, texture, edge sharpness) are optimized based on specific datasets or imaging conditions. While these features may perform well in controlled environments or within the dataset they were designed for. However, they often fail to generalize when applied to new or varied data. The problem of generalization is due to several reasons, like variability in imaging modalities, patient diversity. A model with high performance on multiples dataset, will be a huge advancement in the field of CADx systems.

Traditional machine learning models (like decision trees or logistic regression) offer relatively high interpretability, as their decision-making processes are often transparent. However, modern deep learning models, especially convolutional neural networks (CNNs), are often considered “black boxes” due to their complex internal representations, making it difficult to trace how a diagnosis was reached. This lack of transparency can encumber clinical adoption, reduce trust, and complicate monitoring approval. As a result, interpretability represents another axe of development in this thesis.

## Contributions

- Yermes, M.E.L.A., Debakla, M., Djemal, K. (2024). Mammographic masses descriptor for breast cancer classification and automatic diagnosis. *Revue d'Intelligence Artificielle*, Vol. 38, No. 2, pp. 439-448. <https://doi.org/10.18280/ria.380207>.
- Yermes, M.E.L.A., Debakla, M., Djemal, K. (2021). A CAD System for breast cancer diagnosis using TAR calculation. 9th (Online) International Conference on Applied Analysis and Mathematical Modeling (ICAAMM21) June 11-13, 2021, Istanbul-Turkey. Page 89.
- Enhanced Microcalcification Classification in Mammography Using a Stacking Ensemble of Deep Learning Models; Accepted in MCCSAI'2025 Mediterranean Conference on Computer Science and Artificial Intelligence that will be held from 17<sup>th</sup> to 18<sup>th</sup> of May 2025.

## References

- [1] World Health Organization. (2023). Breast cancer. <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>.
- [2] Xiong, X., Zheng, L.W., Ding, Y. et al. Breast cancer: pathogenesis and treatments. *Sig Transduct Target Ther* 10, 49 (2025). <https://doi.org/10.1038/s41392-024-02108-4>
- [3] Giger, M. L., Karssemeijer, N., & Schnabel, J. A. (2008). Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer. *Annual Review of Biomedical Engineering*, 10, 327–357. <https://doi.org/10.1146/annurev.bioeng.10.061807.160524>
- [4] Elmore, J. G., Wells, C. K., Lee, C. H., Howard, D. H., & Feinstein, A. R. (1994). Variability in radiologists' interpretations of mammograms. *New England Journal of Medicine*, 331(22), 1493–1499. <https://doi.org/10.1056/NEJM199412013312206>
- [5] Roy, S. D., Das, S., Kar, D., Schwenker, F., & Sarkar, R. (2021). Computer Aided Breast Cancer Detection Using Ensembling of Texture and Statistical Image Features. *Sensors*, 21(11), 3628. <https://doi.org/10.3390/s21113628>
- [6] Iqbal, S., N. Qureshi, A., Li, J. et al. On the Analyses of Medical Images Using Traditional Machine Learning Techniques and Convolutional Neural Networks. *Arch Computat Methods Eng* 30, 3173–3233 (2023). <https://doi.org/10.1007/s11831-023-09899-9>
- [7] A. Batool and Y. -C. Byun, "Toward Improving Breast Cancer Classification Using an Adaptive Voting Ensemble Learning Algorithm," in *IEEE Access*, vol. 12, pp. 12869-12882, 2024, doi: 10.1109/ACCESS.2024.3356602.
- [8] M. S. K. Inan, R. Hasan and F. I. Alam, "A Hybrid Probabilistic Ensemble based Extreme Gradient Boosting Approach For Breast Cancer Diagnosis," *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, NV, USA, 2021, pp. 1029-1035, doi: 10.1109/CCWC51732.2021.9376007.
- [9] Davies, E. L. (2016). Breast cancer. *Medicine*, 44(1), 42-46. <https://doi.org/10.1016/j.mpmed.2015.10.002>.
- [10] Abu-Helalah, M., Azab, B., Mubaidin, R. et al. BRCA1 and BRCA2 genes mutations among high-risk breast cancer patients in Jordan. *Sci Rep* 10, 17573 (2020). <https://doi.org/10.1038/s41598-020-74250-2>
- [11] American Cancer Society. (2023). *Breast Cancer Signs and Symptoms*. Retrieved from <https://www.cancer.org/cancer/breast-cancer/about/breast-cancer-signs-and-symptoms.html>
- [12] Moore, K. L., & Dalley, A. F. (2018). *Clinically oriented anatomy* (8th ed.). Wolters Kluwer.
- [13] Ross, M. H., & Pawlina, W. (2020). *Histology: A text and atlas with correlated cell and molecular biology* (8th ed.). Wolters Kluwer

- [14] Zelger, B. (2002). Connective Tissue Tumors. In: Dummer, R., Nestle, F.O., Burg, G. (eds) *Cancers of the Skin. Recent Results in Cancer Research*, vol 160. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-59410-6\\_39](https://doi.org/10.1007/978-3-642-59410-6_39)
- [15] Cunningham, F.G., Leveno, K.J., Bloom, S.L., Spong, C.Y., Dashe, J.S., Hoffman, B.L., Casey, B.M., & Sheffield, J.S., 2018. *Williams Obstetrics*, 24th edition. ed. New York: McGraw-Hill Education/Medical, United States
- [16] S. O'Grady, M.P. Morgan, Microcalcifications in breast cancer: From pathophysiology to diagnosis and prognosis, *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, Volume 1869, Issue 2, 2018, Pages 310-320, ISSN 0304-419X, <https://doi.org/10.1016/j.bbcan.2018.04.006>.
- [17] P. Henrot, A. Leroux, C. Barlier, P. Génin, Breast microcalcifications: The lesions in anatomical pathology, *Journal de Radiologie Diagnostique et Interventionnelle*, Volume 95, Issue 2, 2014, Pages 145-157, ISSN 2211-5706, <https://doi.org/10.1016/j.jradio.2013.12.007>.
- [18] D'Orsi, C. J., Bassett, L. W., Berg, W. A., Feig, S. A., Jackson, V. P., Kopans, D. B., Linver, M. N., Mendelson, E. B., Moss, L. J., and Sickles., E. A. (2003). *American college of radiology (breast imaging reporting and data system). Troisième édition française réalisée par SFR (Société Française de Radiologie)*.
- [19] Berment, H., Becette, V., Mohallem, M., Ferreira, F., & Chérel, P. (2014). Masses in mammography: What are the underlying anatomopathological lesions? *Diagnostic and Interventional Imaging*, 95(2), 124–133.
- [20] F.T.H. Bodewes, A.A. van Asselt, M.D. Dorrius, M.J.W. Greuter, G.H. de Bock, Mammographic breast density and the risk of breast cancer: A systematic review and meta-analysis, *The Breast*, Volume 66, 2022, Pages 62-68, ISSN 0960-9776, <https://doi.org/10.1016/j.breast.2022.09.007>.
- [21] Butler, R., Durand, M., Lanjevar, S., Podany, P., Andrejeva, L., Harigopal, M. (2024). Architectural Distortions. In: Harigopal, M., Andrejeva, L., Singh, K., Lewin, J. (eds) *Radiology Pathology Correlations of Breast Lesions*. Springer, Cham. [https://doi.org/10.1007/978-3-031-65711-5\\_7](https://doi.org/10.1007/978-3-031-65711-5_7)
- [22] de Vita, R., Buccheri, E.M., Villanucci, A. et al. Breast Asymmetry, Classification, and Algorithm of Treatment: Our Experience. *Aesth Plast Surg* 43, 1439–1450 (2019). <https://doi.org/10.1007/s00266-019-01489-0>
- [23] N. Aristokli, I. Polycarpou, S.C. Themistocleous, D. Sophocleous, I. Mamais, Comparison of the diagnostic performance of Magnetic Resonance Imaging (MRI), ultrasound and mammography for detection of breast cancer based on tumor type, breast density and patient's history: A review, *Radiography*, Volume 28, Issue 3, 2022, Pages 848-856, ISSN 1078-8174, <https://doi.org/10.1016/j.radi.2022.01.006>.

- [24] Mohammadi, S., & Ahmadi Livani, M. (2023). A Review of CAD systems for Breast Mass Detection in Mammography Based on Deep Learning. Preprints. <https://doi.org/10.20944/preprints202305.1832.v1>
- [25] Laura Liberman, Jennifer H. Menell, Breast imaging reporting and data system (BI-RADS), Radiologic Clinics of North America, Volume 40, Issue 3, 2002, Pages 409-430, ISSN 0033-8389, [https://doi.org/10.1016/S0033-8389\(01\)00017-3](https://doi.org/10.1016/S0033-8389(01)00017-3).
- [26] Velayuthapandian, K., Karuppiah, G., Vadivel, S. R. S., & Joseph, D. R. V. (2024). Mammogram data analysis: Trends, challenges, and future directions. In D. J. Hemanth (Ed.), *Computational intelligence and modelling techniques for disease detection in mammogram images* (pp. 1–38). Academic Press. <https://doi.org/10.1016/B978-0-443-13999-4.00016-X>
- [27] C. Balleyguier, B. Boyer, A. Athanasiou, D. Vanel, R. Sigal, Comprendre et utiliser le CAD (Aide informatisée au diagnostic) en mammographie, Journal de Radiologie, Volume 86, Issue 1, 2005, Pages 29-35, ISSN 0221-0363, [https://doi.org/10.1016/S0221-0363\(05\)81319-8](https://doi.org/10.1016/S0221-0363(05)81319-8).
- [28] Katzen, J., & Dodelzon, K. (2018). A review of computer aided detection in mammography. *Clinical Imaging*, 52, 305–309. <https://doi.org/10.1016/j.clinimag.2018.08.014>
- [29] George, K. M. J., & Dhas, D. A. S. (2017, March). Preprocessing filters for mammogram images: A review. In *2017 Conference on Emerging Devices and Smart Systems (ICEDSS)* (pp. 1–7). IEEE. <https://doi.org/10.1109/ICEDSS.2017.8073694>
- [30] Mudrakola, S., & Hegde, N. (2023). Removal of noise on mammogram breast images using filtering methods. *Concurrency and Computation: Practice and Experience*, 35(10), e7444. <https://doi.org/10.1002/cpe.7444>
- [31] Pitas, I., Venetsanopoulos, A.N. (1990). Median Filters. In: *Nonlinear Digital Filters. The Springer International Series in Engineering and Computer Science*, vol 84. Springer, Boston, MA. [https://doi.org/10.1007/978-1-4757-6017-0\\_4](https://doi.org/10.1007/978-1-4757-6017-0_4)
- [32] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," in *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961-1005, Sept. 1990, doi: 10.1109/18.57199
- [33] Cleland, T.A. (2009). Contrast Enhancement. In: Binder, M.D., Hirokawa, N., Windhorst, U. (eds) *Encyclopedia of Neuroscience*. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-540-29678-2\\_1250](https://doi.org/10.1007/978-3-540-29678-2_1250)
- [34] Vijayalakshmi, D., Nath, M.K. & Acharya, O.P. A Comprehensive Survey on Image Contrast Enhancement Techniques in Spatial Domain. *Sens Imaging* 21, 40 (2020). <https://doi.org/10.1007/s11220-020-00305-3>
- [35] Härtinger, P., Steger, C. Adaptive histogram equalization in constant time. *J Real-Time Image Proc* 21, 93 (2024). <https://doi.org/10.1007/s11554-024-01465-1>

- [36] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas and K. E. Muller, "Contrast-limited adaptive histogram equalization: speed and effectiveness," *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, Atlanta, GA, USA, 1990, pp. 337-345, doi: 10.1109/VBC.1990.109340.
- [37] Raman B. Paranjape, Chapter 1 - Fundamental Enhancement Techniques, Editor(s): ISAAC N. BANKMAN, *Handbook of Medical Image Processing and Analysis (Second Edition)*, Academic Press, 2009, Pages 3-18, ISBN 9780123739049, <https://doi.org/10.1016/B978-012373904-9.50008-8>.
- [38] Gorgel, P., Sertbas, A. & Ucan, O.N. A Wavelet-Based Mammographic Image Denoising and Enhancement with Homomorphic Filtering. *J Med Syst* 34, 993–1002 (2010). <https://doi.org/10.1007/s10916-009-9316-3>
- [39] Ayyala RS, Chorlton M, Behrman RH, Kornguth PJ, Slanetz PJ. Digital mammographic artifacts on full-field systems: what are they and how do I fix them? *Radiographics*. 2008 Nov-Dec;28(7):1999-2008. doi: 10.1148/rg.287085053. PMID: 19001654.
- [40] Albert, S., Wichtmann, B. D., Zhao, W., Maurer, A., Hesser, J., Attenberger, U. I., Schad, L. R., & Zöllner, F. G. (2023). Comparison of Image Normalization Methods for Multi-Site Deep Learning. *Applied Sciences*, 13(15), 8923. <https://doi.org/10.3390/app13158923>
- [41] H. W. Herwanto, A. N. Handayani, A. P. Wibawa, K. L. Chandrika and K. Arai, "Comparison of Min-Max, Z-Score and Decimal Scaling Normalization for Zoning Feature Extraction on Javanese Character Recognition," *2021 7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, Malang, Indonesia, 2021, pp. 1-3, doi: 10.1109/ICEEIE52663.2021.9616665.
- [42] Ma, J., He, Y., Li, F. et al. Segment anything in medical images. *Nat Commun* 15, 654 (2024). <https://doi.org/10.1038/s41467-024-44824-z>
- [43] Sandra Jardim, João António, Carlos Mora, Image thresholding approaches for medical image segmentation - short literature review, *Procedia Computer Science*, Volume 219, 2023, Pages 1485-1492, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2023.01.439>.
- [44] Lahouaoui LALAOUI and Tayeb MOHAMADI, "A comparative study of Image Region-Based Segmentation Algorithms" *International Journal of Advanced Computer Science and Applications (IJACSA)*, 4(6),2013.<http://dx.doi.org/10.14569/IJACSA.2013.040627>
- [45] H. Aydın and B. Dizdaroğlu, "Medical image segmentation with edge-based level sets: Mobile client-server application," *2017 International Conference on Computer Science and Engineering (UBMK)*, Antalya, Turkey, 2017, pp. 994-999, doi: 10.1109/UBMK.2017.8093412.
- [46] C. Kaushal, A. Singla and P. Panwar, "A Brief Review on Clustering Based Medical Image Segmentation Algorithms with Issues and Challenges," *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future*

- Directions) (ICRITO)*, Noida, India, 2021, pp. 1-8, doi: 10.1109/ICRITO51393.2021.9596154.
- [47] Md. Eshmam Rayed, S.M. Sajibul Islam, Sadia Islam Niha, Jamin Rahman Jim, Md Mohsin Kabir, M.F. Mridha, Deep learning for medical image segmentation: State-of-the-art advancements and challenges, *Informatics in Medicine Unlocked*, Volume 47, 2024, 101504, ISSN 2352-9148, <https://doi.org/10.1016/j.imu.2024.101504>.
- [48] A. Humeau-Heurtier, "Texture Feature Extraction Methods: A Survey," in *IEEE Access*, vol. 7, pp. 8975-9000, 2019, doi: 10.1109/ACCESS.2018.2890743
- [49] Temerik, S.M., Elwahab, S.M.A., Wahman, M.M. et al. Relation between morphological features of initial breast MRI and breast cancer molecular subtypes. *Egypt J Radiol Nucl Med* 54, 147 (2023). <https://doi.org/10.1186/s43055-023-01087-w>
- [50] Kumar, V., Patra, S.K. (2021). Feature Engineering for Machine Learning and Deep Learning Assisted Wireless Communication. In: Oliva, D., Houssein, E.H., Hinojosa, S. (eds) *Metaheuristics in Machine Learning: Theory and Applications*. Studies in Computational Intelligence, vol 967. Springer, Cham. [https://doi.org/10.1007/978-3-030-70542-8\\_4](https://doi.org/10.1007/978-3-030-70542-8_4)
- [51] J. Suto, S. Oniga and P. P. Sitar, "Comparison of wrapper and filter feature selection algorithms on human activity recognition," *2016 6th International Conference on Computers Communications and Control (ICCCC)*, Oradea, Romania, 2016, pp. 124-129, doi: 10.1109/ICCCC.2016.7496749.
- [52] Eduardo F. Morales, Hugo Jair Escalante, Chapter 6 - A brief introduction to supervised, unsupervised, and reinforcement learning, Editor(s): Alejandro A. Torres-García, Carlos A. Reyes-García, Luis Villaseñor-Pineda, Omar Mendoza-Montoya, *Biosignal Processing and Classification Using Computational Learning and Intelligence*, Academic Press, 2022, Pages 111-129, ISBN 9780128201251, <https://doi.org/10.1016/B978-0-12-820125-1.00017-8>.
- [53] Saad Almutairi, Manimurugan S., Byung-Gyu Kim, Majed M. Aborokbah, Narmatha C., Breast cancer classification using Deep Q Learning (DQL) and gorilla troops optimization (GTO), *Applied Soft Computing*, Volume 142, 2023, 110292, ISSN 1568-4946, <https://doi.org/10.1016/j.asoc.2023.110292>.
- [54] Omar Adil Deheyab, A., Alwan, M. H., Abdul Rezzaqe, I. K., Mahmood, O. A., Hammadi, Y. I., Kareem, A. N., & Ibrahim, M. (2022). *An overview of challenges in medical image processing*. In *Proceedings of the 6th International Conference on Future Networks & Distributed Systems (ICFNDS '22)* (pp. 511–516). ACM. <https://doi.org/10.1145/3584202.3584278>
- [55] Long Gao, Lei Zhang, Chang Liu, Shandong Wu, Handling imbalanced medical image data: A deep-learning-based one-class classification approach, *Artificial Intelligence in*

Medicine, Volume 108, 2020, 101935, ISSN 0933-3657, <https://doi.org/10.1016/j.artmed.2020.101935>.

- [56] A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, *Inf. Fusion* 58 (2020) 82–115.
- [57] A.J. Barnett, F.R. Schwartz, C. Tao, C. Chen, Y. Ren, J.Y. Lo, C. Rudin, A case-based interpretable deep learning model for classification of mass lesions in digital mammography, *Nat. Mach. Intell.* 3 (12) (2021) 1061–1070.
- [58] Jun Gao, Qian Jiang, Bo Zhou, Daozheng Chen. Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview[J]. *Mathematical Biosciences and Engineering*, 2019, 16(6): 6536-6561. doi: 10.3934/mbe.2019326
- [59] Petrick, N., Sahiner, B., Armato, S. G., 3rd, Bert, A., Correale, L., Delsanto, S., Freedman, M. T., Fryd, D., Gur, D., Hadjiiski, L., Huo, Z., Jiang, Y., Morra, L., Paquerault, S., Raykar, V., Samuelson, F., Summers, R. M., Tourassi, G., Yoshida, H., Zheng, B., ... Chan, H. P. (2013). Evaluation of computer-aided detection and diagnosis systems. *Medical physics*, 40(8), 087001. <https://doi.org/10.1118/1.4816310>
- [60] Qiu, J. (2024). An Analysis of Model Evaluation with Cross-Validation: Techniques, Applications, and Recent Advances. *Advances in Economics, Management and Political Sciences*, 99, 69-72.
- [61] Narváez, F., Díaz, G., Romero, E. (2010). Automatic BI-RADS Description of Mammographic Masses. In: Martí, J., Oliver, A., Freixenet, J., Martí, R. (eds) *Digital Mammography. IWDM 2010. Lecture Notes in Computer Science*, vol 6136. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-13666-5\\_91GHG](https://doi.org/10.1007/978-3-642-13666-5_91GHG)
- [62] Lee, Christoph I., and others, 'Spiculated Masses', in Christoph I. Lee, Constance D. Lehman, and Lawrence W. Bassett (eds), *Breast Imaging*, Rotations in Radiology (New York, 2018; online edn, Oxford Academic, 1 Mar. 2018), <https://doi.org/10.1093/med/9780190270261.003.0025>
- [63] Reinhard Klette, Joviša Žunić, ADR shape descriptor – Distance between shape centroids versus shape diameter, *Computer Vision and Image Understanding*, Volume 116, Issue 6, 2012, Pages 690-697, ISSN 1077-3142, <https://doi.org/10.1016/j.cviu.2012.02.001>.
- [64] Shape Measurement - an overview | ScienceDirect Topics. (n.d.). Retrieved April 13, 2025, from <https://www.sciencedirect.com/topics/computer-science/shape-measurement>
- [65] Zheng, C., & Sun, D. W. (2008). Object Measurement Methods. *Computer Vision Technology for Food Quality Evaluation*, 57–80. <https://doi.org/10.1016/B978-012373642-0.50006-5>

- [66] Rosin PL (2003) Measuring shape: ellipticity, rectangularity, and triangularity. *Mach vis Appl* 14:172–184. <https://doi.org/10.1007/s00138-002-0118-6>
- [67] D. Žunić, C. Martinez-Ortiz, J. J. I. J. o. P. R. ŽUNIĆ and A. Intelligence, *Shape rectangularity measures*, vol. 26, no. 06, pp. 1254002, 2012.
- [68] Burger, W., Burge, M.J. (2013). Fourier Shape Descriptors. In: *Principles of Digital Image Processing. Undergraduate Topics in Computer Science*. Springer, London. [https://doi.org/10.1007/978-1-84882-919-0\\_6](https://doi.org/10.1007/978-1-84882-919-0_6)
- [69] M. E. Celebi and Y. A. Aslandogan, "A comparative study of three moment-based shape descriptors," *International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume II*, Las Vegas, NV, USA, 2005, pp. 788-793 Vol. 1, doi: 10.1109/ITCC.2005.3.
- [70] Chiang, A. & Liao, S. (2015). Image Analysis with Legendre Moment Descriptors. *Journal of Computer Science*, 11(1), 127-136. <https://doi.org/10.3844/jcssp.2015.127.136>
- [71] Zhu, W., Lu, X., Xu, T., & Zhao, Z. (2014). *High order structure descriptors for scene images*. arXiv. <https://arxiv.org/abs/1410.3910>
- [72] B. B. Mandelbrot, "The Fractal Geometry of Nature," W.H. Freeman and Company, New York, 1983.
- [73] Albertovich, T. D., & Aleksandrovna, R. I. (2017). The Fractal Analysis of the Images and Signals in Medical Diagnostics. InTech. doi: 10.5772/intechopen.68167
- [74] M. Y. Marusina and E. A. Karaseva, "Automatic Analysis of Medical Images Based on Fractal Methods," *2019 International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS)*, Sochi, Russia, 2019, pp. 349-352, doi: 10.1109/ITQMIS.2019.8928378.
- [75] Kane, L., Khanna, P. Real-time recognition of medial structures within hand postures through Eigen-space and geometric skeletal shape features. *Multimed Tools Appl* 76, 4571–4598 (2017). <https://doi.org/10.1007/s11042-016-4173-9>
- [76] Vinit Vijay Deshpande, Romana Piat, A skeletonization based image segmentation algorithm to isolate slender regions in 3D microstructures, *Materials & Design*, Volume 252, 2025, 113765, ISSN 0264-1275, <https://doi.org/10.1016/j.matdes.2025.113765>.
- [77] Leslie Verscheure, Laurent Peyrodie, Anne- Sophie Dewalle, Nicolas Reys, Nacim Betrouni, et al.. Three-dimensional skeletonization and symbolic description in vascular imaging: preliminary results. *International Journal of Computer Assisted Radiology and Surgery*, 2013, 8 (2), pp.233-246. 10.1007/s11548-012-0784-4
- [78] Anne Humeau-Heurtier. Texture feature extraction methods: A survey. *IEEE Access*, 2019, 7, pp.8975-9000.10.1109/ACCESS.2018.2890743.
- [79] Armi, L., & Fekri-Ershad, S. (2019). *Texture image analysis and texture classification methods – A review*. arXiv. <https://arxiv.org/abs/1904.06554>

- [80] Erchan Aptoula, Sébastien Lefèvre, Chapter 1 - Morphological Texture Description of Grey-Scale and Color Images, Editor(s): Peter W. Hawkes, *Advances in Imaging and Electron Physics*, Elsevier, Volume 169, 2011, Pages 1-74, ISSN 1076-5670, ISBN 9780123859815, <https://doi.org/10.1016/B978-0-12-385981-5.00001-X>.
- [81] Tomita, F., Tsuji, S. (1990). Statistical Texture Analysis. In: *Computer Analysis of Visual Textures*. The Springer International Series in Engineering and Computer Science, vol 102. Springer, Boston, MA. [https://doi.org/10.1007/978-1-4613-1553-7\\_2](https://doi.org/10.1007/978-1-4613-1553-7_2)
- [82] A. Ramola, A. K. Shakya, and D. V. Pham, "Study of Statistical Methods for Texture Analysis and Their Modern Evolutions", *Eng. Reports*, Vol. 2, No. 4: e12149, pp. 1–24, 2020, doi:10.1002/eng2.12149.
- [83] Aggarwal, N. and K. Agrawal, R. (2012) First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images. *Journal of Signal and Information Processing*, 3, 146-153. doi: 10.4236/jsip.2012.32019.
- [84] Martin Hassner, Jack Sklansky, The use of Markov Random Fields as models of texture, *Computer Graphics and Image Processing*, Volume 12, Issue 4, 1980, Pages 357-370, ISSN 0146-664X, [https://doi.org/10.1016/0146-664X\(80\)90019-2](https://doi.org/10.1016/0146-664X(80)90019-2).
- [85] H. Mhidra, J. Brochard, M. Leard, AR models and bidimensional discrete moments applied to texture modelling and recognition, *Pattern Recognition*, Volume 26, Issue 5, 1993, Pages 721-726, ISSN 0031-3203, [https://doi.org/10.1016/0031-3203\(93\)90124-F](https://doi.org/10.1016/0031-3203(93)90124-F).
- [86] M. M. Islam, D. Zhang and G. Lu, "Rotation invariant curvelet features for texture image retrieval," *2009 IEEE International Conference on Multimedia and Expo*, New York, NY, USA, 2009, pp. 562-565, doi: 10.1109/ICME.2009.5202558.
- [87] S Arivazhagan, L Ganesan, Texture classification using wavelet transform, *Pattern Recognition Letters*, Volume 24, Issues 9–10, 2003, Pages 1513-1521, ISSN 0167-8655, [https://doi.org/10.1016/S0167-8655\(02\)00390-2](https://doi.org/10.1016/S0167-8655(02)00390-2).
- [88] A. A. Ursani, K. Kpalma and J. Ronsin, "Texture features based on Fourier transform and Gabor filters: an empirical comparison," *2007 International Conference on Machine Vision*, Isalambad, Pakistan, 2007, pp. 67-72, doi: 10.1109/ICMV.2007.4469275.
- [89] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615-1630, Oct. 2005, doi: 10.1109/TPAMI.2005.188.
- [90] Amanatiadis, A., Kaburlasos, V. G., Gasteratos, A., & Papadakis, S. E. (2011). Evaluation of shape descriptors for shape-based image retrieval. *IET Image Processing*, 5(5), 493–499. <https://doi.org/10.1049/iet-ipr.2009.0246>
- [91] World Health Organization International, Agency of Research on Caner: PRESS RELEASE N° 292, 15 December 2020.

- [92] Singh, L., Jaffery, Z.A. (2018). Computerized detection of breast cancer in digital mammograms. *International Journal of Computers and Applications*, 40(2): 98-109. <https://doi.org/10.1080/1206212X.2017.1395131>
- [93] Nass, S., Henderson, I., Lashof, J. (2001). Committee on technologies for the early detection of breast cancer. In *Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer*. Nat. Cancer Policy Board, Inst. Med., Commission Life Stud, Nat. Res. Council.
- [94] Burnside, E.S., Sickles, E.A., Bassett, L.W., Rubin, D.L., Lee, C.H., Ikeda, D.M., Mendelson, E.B., Wilcox, P.A., Butler, P.F., D'Orsi, C.J. (2009). The ACR BI-RADS® experience: Learning from history. *Journal of the American College of Radiology*, 6(12): 851-860. <https://doi.org/10.1016/j.jacr.2009.07.023>
- [95] Berment, H., Becette, V., Mohallem, M., Ferreira, F., Chérel, P. (2014). Masses in mammography: What are the underlying anatomopathological lesions?. *Diagnostic and Interventional Imaging*, 95(2): 124-133. <https://doi.org/10.1016/j.diii.2013.12.010>
- [96] Madhukar, B.N., Bharathi, S.H., Polnaya, A.M. (2023). Multi-scale convolution based breast cancer image segmentation with attention mechanism in conjunction with war search optimization. *International Journal of Computers and Applications*, 45(5): 353-366. <https://doi.org/10.1080/1206212X.2023.2212945>
- [97] Kaushal, C., Bhat, S., Koundal, D., Singla, A. (2019). Recent trends in computer assisted diagnosis (CAD) system for breast cancer diagnosis using histopathological images. *IRBM*, 40(4): 211-227. <https://doi.org/10.1016/j.irbm.2019.06.001>
- [98] Katzen, J., Dodelzon, K. (2018). A review of computer aided detection in mammography. *Clinical Imaging*, 52: 305-309. <https://doi.org/10.1016/j.clinimag.2018.08.014>
- [99] Guliato, D., Rangayyan, R.M., de Carvalho, J.D., Santiago, S.A. (2006). Spiculation-preserving polygonal modeling of contours of breast tumors. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, USA*, pp. 2791-2794. <https://doi.org/10.1109/IEMBS.2006.260441>
- [100] Xie, W., Li, Y., Ma, Y. (2016). Breast mass classification in digital mammography based on extreme learning machine. *Neurocomputing*, 173: 930-941. <https://doi.org/10.1016/j.neucom.2015.08.048>
- [101] Beheshti, S.M.A., Noubari, H.A., Fatemizadeh, E., Khalili, M. (2016). Classification of abnormalities in mammograms by new asymmetric fractal features. *Biocybernetics and Biomedical Engineering*, 36(1): 56-65. <https://doi.org/10.1016/j.bbe.2015.07.002>
- [102] Souza, J.C., Silva, T.F.B., Rocha, S.V., Paiva, A.C., Braz, G., Almeida, J.D.S., Silva, A.C. (2017). Classification of malignant and benign tissues in mammography using dental shape descriptors and shape distribution. *7th Latin American Conference on Networked and Electronic Media (LACNEM 2017)*, pp. 22-27. <https://doi.org/10.1049/ic.2017.0030>

- [103] Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D. (2002). Shape distributions. *ACM Transactions on Graphics (TOG)*, 21(4): 807-832. <https://doi.org/10.1145/571647.571648>
- [104] Ribli, D., Horváth, A., Unger, Z., Pollner, P., Csabai, I. (2018). Detecting and classifying lesions in mammograms with deep learning. *Scientific Reports*, 8(1): 4165. <https://doi.org/10.1038/s41598-018-22437-z>
- [105] Goudarzi, M., Maghooli, K. (2018). Extraction of fuzzy rules at different concept levels related to image features of mammography for diagnosis of breast cancer. *Biocybernetics and Biomedical Engineering*, 38(4): 1004-1014. <https://doi.org/10.1016/j.bbe.2018.09.002>
- [106] Sun, L., Wang, J., Hu, Z., Xu, Y., Cui, Z. (2019). Multi-view convolutional neural networks for mammographic image classification. *IEEE Access*, 7: 126273-126282. <https://doi.org/10.1109/ACCESS.2019.2939167>
- [107] Vijayarajeswari, R., Parthasarathy, P., Vivekanandan, S., Basha, A.A. (2019). Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform. *Measurement*, 146: 800-805. <https://doi.org/10.1016/j.measurement.2019.05.083>
- [108] Pezeshki, H., Rastgarpour, M., Sharifi, A., Yazdani, S. (2019). Extraction of spiculated parts of mammogram tumors to improve accuracy of classification. *Multimedia Tools and Applications*, 78: 19979-20003. <https://doi.org/10.1007/s11042-019-7185-4>
- [109] de Brito Silva, T.F., de Paiva, A.C., Silva, A.C., Braz Júnior, G., de Almeida, J.D.S. (2020). Classification of breast masses in mammograms using geometric and topological feature maps and shape distribution. *Research on Biomedical Engineering*, 36: 225-235. <https://doi.org/10.1007/s42600-020-00063-x>
- [110] Arora, R., Rai, P.K., Raman, B. (2020). Deep feature-based automatic classification of mammograms. *Medical & Biological Engineering & computing*, 58: 1199-1211. <https://doi.org/10.1007/s11517-020-02150-8>
- [111] Zhang, Y.D., Satapathy, S.C., Guttery, D.S., Górriz, J.M., Wang, S.H. (2021). Improved breast cancer classification through combining graph convolutional network and convolutional neural network. *Information Processing & Management*, 58(2): 102439. <https://doi.org/10.1016/j.ipm.2020.102439>
- [112] Kohli, A., Jha, S. (2018). Why CAD failed in mammography. *Journal of the American College of Radiology*, 15(3 Pt B): 535-537. <https://doi.org/10.1016/j.jacr.2017.12.029>
- [113] Douglas, D.H., Peucker, T.K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2): 112-122. <https://doi.org/10.3138/FM57-6770-U75U-7727>
- [114] Alajlan, N., El Rube, I., Kamel, M.S., Freeman, G. (2007). Shape retrieval using triangle-area representation and dynamic space warping. *Pattern Recognition*, 40(7):

- 1911-1920. <https://doi.org/10.1016/j.patcog.2006.12.005>
- [115] Bezdek, J.C., Ehrlich, R., Full, W. (1984). FCM: The fuzzy C-means clustering algorithm. *Computers & Geosciences*, 10(2-3): 191-203. [https://doi.org/10.1016/0098-3004\(84\)90020-7](https://doi.org/10.1016/0098-3004(84)90020-7)
- [116] Lee, R. S., Gimenez, F., Hoogi, A. Rubin, D. L. The Cancer Imaging Archive. <http://dx.doi.org/10.7937/K9/TCIA.2016.7O02S9CY>
- [117] Rabidas R, Midya A, Chakraborty J. Neighborhood structural Similarity mapping for the classification of masses in mammograms. *IEEE J. Biomed. Health Inf.* 2018; 22(3):826–34.
- [118] P. Kaur, G. Singh, P. Kaur, Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification, *Inf. Med. Unlocked* 100239 (2019), <https://doi.org/10.1016/j.imu.2019.100239>.
- [119] Andrés Anaya-Isaza, Leonel Mera-Jiménez, Martha Zequera-Diaz, An overview of deep learning in medical imaging, *Informatics in Medicine Unlocked*, Volume 26, 2021, 100723, ISSN 2352-9148, <https://doi.org/10.1016/j.imu.2021.100723>.
- [120] Aggarwal, R., Sounderajah, V., Martin, G. et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *npj Digit. Med.* 4, 65 (2021). <https://doi.org/10.1038/s41746-021-00438-z>
- [121] Shanthi H J, Hamza Mohammed Ridha Al-Khafaji, Salwan Mohammed Shaheed, Rajasekhar Pittala, Harendra Singh Negi, Arun Pratap Srivastava, Navneet Kumar, Anurag Shrivastava. (2024). Deep Learning Applications in Medical Image Analysis: U-Net for Diagnosis. *International Journal of Intelligent Systems and Applications in Engineering*, 12(21s), 494–500. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/5446>
- [122] Toosi, A., Bottino, A. G., Saboury, B., Siegel, E., & Rahmim, A. (2021). A brief history of AI: How to prevent another winter (A critical review). *PET Clinics*, 16(4), 449–469. <https://doi.org/10.1016/j.cpet.2021.07.001>
- [123] Mohsen Soori, Behrooz Arezoo, Roza Dastres, Artificial intelligence, machine learning and deep learning in advanced robotics, a review, *Cognitive Robotics*, Volume 3, 2023, Pages 54-70, ISSN 2667-2413, <https://doi.org/10.1016/j.cogr.2023.04.001>.
- [124] Khemani, B., Patil, S., Kotecha, K. et al. A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions. *J Big Data* 11, 18 (2024). <https://doi.org/10.1186/s40537-023-00876-4>
- [125] D’Addona, D.M. (2014). Neural Network. In: Laperrière, L., Reinhart, G. (eds) *CIRP Encyclopedia of Production Engineering*. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-20617-7\\_6563](https://doi.org/10.1007/978-3-642-20617-7_6563)
- [126] Mohaiminul Islam, Guorong Chen, Shangzhu Jin. (2019). An Overview of Neural

- Network. *American Journal of Neural Networks and Applications*, 5(1), 7-11. <https://doi.org/10.11648/j.ajjna.20190501.12>.
- [127] O. K. Oyedotun, A. E. R. Shabayek, D. Aouada and B. Ottersten, "Going Deeper With Neural Networks Without Skip Connections," *2020 IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates, 2020, pp. 1756-1760, doi: 10.1109/ICIP40778.2020.9191356.
- [128] B. Ding, H. Qian and J. Zhou, "Activation functions and their characteristics in deep neural networks," *2018 Chinese Control And Decision Conference (CCDC)*, Shenyang, China, 2018, pp. 1836-1841, doi: 10.1109/CCDC.2018.8407425.
- [129] R. Parhi and R. D. Nowak, "The Role of Neural Network Activation Functions," in *IEEE Signal Processing Letters*, vol. 27, pp. 1779-1783, 2020, doi: 10.1109/LSP.2020.3027517.
- [130] Andrea Apicella, Francesco Donnarumma, Francesco Isgrò, Roberto Prevete, A survey on modern trainable activation functions, *Neural Networks*, Volume 138, 2021, Pages 14-32, ISSN 0893-6080, <https://doi.org/10.1016/j.neunet.2021.01.026>.
- [131] Adamantios Zaras, Nikolaos Passalis, Anastasios Tefas, Chapter 2 - Neural networks and backpropagation, Editor(s): Alexandros Iosifidis, Anastasios Tefas, *Deep Learning for Robot Perception and Cognition*, Academic Press, 2022, Pages 17-34, ISBN 9780323857871, <https://doi.org/10.1016/B978-0-32-385787-1.00007-5>.
- [132] Hoss Belyadi, Alireza Haghighat, Chapter 6 - Neural networks and Deep Learning, Editor(s): Hoss Belyadi, Alireza Haghighat, *Machine Learning Guide for Oil and Gas Using Python*, Gulf Professional Publishing, 2021, Pages 297-347, ISBN 9780128219294, <https://doi.org/10.1016/B978-0-12-821929-4.00008-1>.
- [133] M.W Gardner, S.R Dorling, Artificial neural networks (the multilayer perceptron) a review of applications in the atmospheric sciences, *Atmospheric Environment*, Volume 32, Issues 14–15, 1998, Pages 2627-2636, ISSN 1352-2310, [https://doi.org/10.1016/S1352-2310\(97\)00447-0](https://doi.org/10.1016/S1352-2310(97)00447-0).
- [134] Kruse, R., Mostaghim, S., Borgelt, C., Braune, C., Steinbrecher, M. (2022). Multi-layer Perceptrons. In: *Computational Intelligence. Texts in Computer Science*. Springer, Cham. [https://doi.org/10.1007/978-3-030-42227-1\\_5](https://doi.org/10.1007/978-3-030-42227-1_5)
- [135] Zhang, Z., Feng, F., & Huang, T. (2022). FNNS: An Effective Feedforward Neural Network Scheme with Random Weights for Processing Large-Scale Datasets. *Applied Sciences*, 12(23), 12478. <https://doi.org/10.3390/app12231247>
- [136] Saastamoinen, A., Lehtokangas, M., Värri, A., Saarinen, J. (2001). Biomedical Applications of Radial Basis Function Networks. In: Howlett, R.J., Jain, L.C. (eds) *Radial Basis Function Networks 2. Studies in Fuzziness and Soft Computing*, vol 67. Physica, Heidelberg. [https://doi.org/10.1007/978-3-7908-1826-0\\_7](https://doi.org/10.1007/978-3-7908-1826-0_7)
- [137] Marhon, S.A., Cameron, C.J.F., Kremer, S.C. (2013). Recurrent Neural Networks. In: Bianchini, M., Maggini, M., Jain, L. (eds) *Handbook on Neural Information Processing*.

Intelligent Systems Reference Library, vol 49. Springer, Berlin, Heidelberg.  
[https://doi.org/10.1007/978-3-642-36657-4\\_2](https://doi.org/10.1007/978-3-642-36657-4_2)

- [138] Sepp Hochreiter; Jürgen Schmidhuber (1997). "Long short-term memory". *Neural Computation*. 9 (8): 1735-1780. doi:10.1162/neco.1997.9.8.1735. PMID 9377276. S2CID 1915014.
- [139] Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 9, 611–629 (2018). <https://doi.org/10.1007/s13244-018-0639-9>
- [140] Sakshi Indolia, Anil Kumar Goswami, S.P. Mishra, Pooja Asopa, Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach, *Procedia Computer Science*, Volume 132, 2018, Pages 679-688, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2018.05.069>.
- [141] Carriero, A.; Groenhoff, L.; Vologina, E.; Basile, P.; Albera, M. Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024. *Diagnostics* 2024, 14, 848. <https://doi.org/10.3390/diagnostics14080848>
- [142] Watanabe, A.T.; Lim, V.; Vu, H.X.; Chim, R.; Weise, E.; Liu, J.; Bradley, W.G.; Comstock, C.E. Improved Cancer Detection Using Artificial Intelligence: A Retrospective Evaluation of Missed Cancers on Mammography. *J. Digit. Imaging* 2019, 32, 625–637.
- [143] Schaffter, T.; Buist, D.S.M.; Lee, C.I.; Nikulin, Y.; Ribli, D.; Guan, Y.; Lotter, W.; Jie, Z.; Du, H.; Wang, S.; et al. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. *JAMA Netw. Open* 2020, 3, e200265.
- [144] Kim, H.-E.; Kim, H.H.; Han, B.-K.; Kim, K.H.; Han, K.; Nam, H.; Lee, E.H.; Kim, E.-K. Changes in Cancer Detection and False-Positive Recall in Mammography Using Artificial Intelligence: A Retrospective, Multireader Study. *Lancet Digit. Health* 2020, 2, e138–e148.
- [145] Dembrower, K.; Crippa, A.; Colón, E.; Eklund, M.; Strand, F. Artificial Intelligence for Breast Cancer Detection in Screening Mammography in Sweden: A Prospective, Population-Based, Paired-Reader, Non-Inferiority Study. *Lancet Digit. Health* 2023, 5, e703–e711.
- [146] Ng, A.Y.; Oberije, C.J.G.; Ambrózay, É.; Szabó, E.; Serf"oz"o, O.; Karpati, E.; Fox, G.; Glocker, B.; Morris, E.A.; Forrai, G.; et al. Prospective Implementation of AI-Assisted Screen Reading to Improve Early Detection of Breast Cancer. *Nat. Med.* 2023, 29, 3044–3049.
- [147] Romero-Martín, S.; Elías-Cabot, E.; Raya-Povedano, J.L.; Gubern-Mérida, A.; Rodríguez-Ruiz, A.; Álvarez-Benito, M. Stand-Alone Use of Artificial Intelligence for Digital Mammography and Digital Breast Tomosynthesis Screening: A Retrospective Evaluation. *Radiology* 2022, 302, 535–542.

- [148] Zheng, T.; Lin, F.; Li, X.; Chu, T.; Gao, J.; Zhang, S.; Li, Z.; Gu, Y.; Wang, S.; Zhao, F.; et al. Deep Learning-Enabled Fully Automated Pipeline System for Segmentation and Classification of Single-Mass Breast Lesions Using Contrast-Enhanced Mammography: A Prospective, Multicentre Study. *eClinicalMedicine* 2023, 58, 101913
- [149] Beuque, M.P.L.; Lobbes, M.B.I.; Van Wijk, Y.; Widaatalla, Y.; Primakov, S.; Majer, M.; Balleyguier, C.; Woodruff, H.C.; Lambin, P. Combining Deep Learning and Handcrafted Radiomics for Classification of Suspicious Lesions on Contrast-Enhanced Mammograms. *Radiology* 2023, 307, e221843.
- [150] Giang Ngo, Rodney Beard, Rohitash Chandra, Evolutionary bagging for ensemble learning, *Neurocomputing*, Volume 510, 2022, Pages 1-14, ISSN 0925-2312, <https://doi.org/10.1016/j.neucom.2022.08.055>.
- [151] Martinez, R., & Gupta, S. (2022). Breast microcalcifications: Past, present and future. *Molecular and Clinical Oncology*, 16(2), Article 25. Retrieved from <https://pmc.ncbi.nlm.nih.gov/articles/PMC8892454>
- [152] Ali MCK, Hall P, Humphreys K. Association of microcalcification clusters with short-term invasive breast cancer risk and breast cancer risk factors. *Sci Rep*. 2019;9:1–8.
- [153] D. Müller, I. Soto-Rey and F. Kramer, "An Analysis on Ensemble Learning Optimized Medical Image Classification With Deep Convolutional Neural Networks," in *IEEE Access*, vol. 10, pp. 66467-66480, 2022, doi: 10.1109/ACCESS.2022.3182399.
- [154] Ammar Mohammed, Rania Kora, A comprehensive review on ensemble deep learning: Opportunities and challenges, *Journal of King Saud University - Computer and Information Sciences*, Volume 35, Issue 2, 2023, Pages 757-774, ISSN 1319-1578, <https://doi.org/10.1016/j.jksuci.2023.01.014>
- [155] Chatzimparmpas, A., Martins, R., Kucher, K., & Kerren, A. (2020). StackGenVis: Alignment of Data, Algorithms, and Models for Stacking Ensemble Learning Using Performance Metrics. *IEEE Transactions on Visualization and Computer Graphics*, 27, 1547- 1557. <https://doi.org/10.1109/TVCG.2020.3030352>.
- [156] Cai, G., Guo, Y., Chen, W., Zeng, H., Zhou, Y., & Lu, Y. (2020). Computer-aided detection and diagnosis of microcalcification clusters on full field digital mammograms based on deep learning method using neutrosophic boosting. *Multimedia Tools and Applications*, 79(23), 17147–17167. <https://doi.org/10.1007/s11042-019-7726-x>
- [157] Kang, D., Gweon, H.M., Eun, N.L. et al. A convolutional deep learning model for improving mammographic breast-microcalcification diagnosis. *Sci Rep* 11, 23925 (2021). <https://doi.org/10.1038/s41598-021-03516-0>
- [158] A. Gerbasi, G. Clementi, F. Corsi, S. Albasini, A. Malovini, S. Quaglini, R. Bellazzi, DeepMiCa: Automatic segmentation and classification of breast MicroCalcifications from mammograms, *Computer Methods and Programs in Biomedicine*, Volume 235, 2023, 107483, ISSN 0169-2607, <https://doi.org/10.1016/j.cmpb.2023.107483>.
- [159] Tsai, H.-H., Wei, C.-S., Hsieh, Y.-C., Chen, I.-M., Yeh, P.-Y., Shih, D., & Chin, C.-L.

- (2022). Calcification clusters and lesions analysis in mammogram using multi-architecture deep learning algorithms. *Biomedical Engineering: Applications, Basis and Communications*, 34(2), 2250022.
- [160] Teoh JR, Hasikin K, Lai KW, Wu X, Li C. Enhancing early breast cancer diagnosis through automated microcalcification detection using an optimized ensemble deep learning framework. *PeerJ Comput Sci.* 2024 May 29;10:e2082. doi: 10.7717/peerj-cs.2082. PMID: 38855257; PMCID: PMC11157616.
- [161] H. Du, M. M. -S. Yao, S. Liu, L. Chen, W. P. Chan and M. Feng, "Automatic Calcification Morphology and Distribution Classification for Breast Mammograms With Multi-Task Graph Convolutional Neural Network," in *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 8, pp. 3782-3793, Aug. 2023, doi: 10.1109/JBHI.2023.3249404.
- [162] A. Yurdusev, K. Adem, M. Hekim, Detection and classification of microcalcifications in mammograms images using difference filter and Yolov4 deep learning model, *Biomedical Signal Processing and Control*, Volume 80, Part 2, 2023, 104360, ISSN 1746-8094, <https://doi.org/10.1016/j.bspc.2022.104360>
- [163] Kumar Singh, K., Kumar, S., Antonakakis, M., Moirogiorgou, K., Deep, A., Kashyap, K. L., Bajpai, M. K., & Zervakis, M. (2022). Deep Learning Capabilities for the Categorization of Microcalcification. *International Journal of Environmental Research and Public Health*, 19(4), 2159. <https://doi.org/10.3390/ijerph19042159>
- [164] Savelli, B., Bria, A., Molinara, M., Marrocco, C., & Tortorella, F. (2020). A multi-context CNN ensemble for small lesion detection. *Artificial intelligence in medicine*, 103, 101749 .<https://doi.org/10.1016/j.artmed.2019.101749>.
- [165] Curated Breast Imaging Subset of Digital Database for Screening Mammography. <http://doi.org/10.7937/K9/TCIA.2016.7O02S9CY>
- [166] Liu, H., Chen, Y., Zhang, Y., Wang, L., Luo, R., Wu, H., Wu, C., Zhang, H., Tan, W., Yin, H., & Wang, D. (2021). A deep learning model integrating mammography and clinical factors facilitates the malignancy prediction of BI-RADS 4 microcalcifications in breast cancer screening. *European Radiology*, 31, 5902 - 5912. <https://doi.org/10.1007/s00330-020-07659-y>.
- [167] Buades, A., Coll, B., & Morel, J. (2011). Non-Local Means Denoising. *Image Process. Line*, 1. [https://doi.org/10.5201/IPOL.2011.BCM\\_NLM](https://doi.org/10.5201/IPOL.2011.BCM_NLM).
- [168] Alshamrani, K., Alshamrani, H., Alqahtani, F., & Almutairi, B. (2022). Enhancement of Mammographic Images Using Histogram-Based Techniques for Their Classification Using CNN. *Sensors (Basel, Switzerland)*, 23. <https://doi.org/10.3390/s23010235>.
- [169] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700-4708.
- [170] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional

Neural Networks. arXiv preprint arXiv:1905.11946.

- [171] Kumar Singh, K.; Kumar, S.; Antonakakis, M.; Moirogiorgou, K.; Deep, A.; Kashyap, .L.; Bajpai, M.K.; Zervakis, M. Deep Learning Capabilities for the Categorization of Microcalcification. *Int. J. Environ. Res. Public Health* 2022, 19, 2159. <https://doi.org/10.3390/ijerph19042159>
- [172] Rehman, K.u.; Li, J.; Pei, Y.; Yasin, A.; Ali, S.; Mahmood, T. Computer Vision-Based Microcalcification Detection in Digital Mammograms Using Fully Connected Depthwise Separable Convolutional Neural Network. *Sensors* 2021, 21, 4854. <https://doi.org/10.3390/s21144854>
- [173] Chen, Y.; Zhang, Q.; Wu, Y.; Liu, B.; Wang, M.; Lin, Y. Fine-tuning ResNet for breast cancer classification from mammography. In *The International Conference on Healthcare Science and Engineering*; Springer: Singapore, 2018; pp. 83–96.
- [174] Chen, Y.; Zhang, Q.; Wu, Y.; Liu, B.; Wang, M.; Lin, Y. Fine-tuning ResNet for breast cancer classification from mammography. In *The International Conference on Healthcare Science and Engineering*; Springer: Singapore, 2018; pp. 83–96.
- [175] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778.
- [176] Yermes, M.E.L.A., Debakla, M., Djemal, K. (2024). Mammographic masses descriptor for breast cancer classification and automatic diagnosis. *Revue d'Intelligence Artificielle*, Vol. 38, No. 2, pp. 439-448. <https://doi.org/10.18280/ria.380207>.
- [177] Yermes, M.E.L.A., Debakla, M., Djemal, K. (2021). A CAD System for breast cancer diagnosis using TAR calculation. 9th (Online) International Conference on Applied Analysis and Mathematical Modeling (ICAAMM21) June 11-13, 2021, Istanbul-Turkey. Page 89.
- [178] Manojee Roy, Ujwala Baruah, Enhancing medical image classification through controlled diversity in ensemble learning, *Engineering Applications of Artificial Intelligence*, Volume 133, Part B, 2024, 108138, ISSN 0952-1976, <https://doi.org/10.1016/j.engappai.2024.108138>.
- [179] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119–139. <https://doi.org/10.1006/jcss.1997.1504>
- [180] Friedman, J. H. (1999, March). *Stochastic gradient boosting* [Technical report]. Stanford University.
- [181] Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, 785–794. ACM. <https://doi.org/10.1145/2939672.2939785>
- [182] Z. Wang, H. Ren, R. Lu and L. Huang, "Stacking Based LightGBM-CatBoost-

RandomForest Algorithm and Its Application in Big Data Modeling," *2022 4th International Conference on Data-driven Optimization of Complex Systems (DOCS)*, Chengdu, China, 2022, pp. 1-6, doi: 10.1109/DOCS55193.2022.9967714.

- [183] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," in *IEEE Access*, vol. 10, pp. 99129-99149, 2022, doi: 10.1109/ACCESS.2022.3207287.
- [184] Zhou, ZH. (2009). Ensemble Learning. In: Li, S.Z., Jain, A. (eds) *Encyclopedia of Biometrics*. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-73003-5\\_293](https://doi.org/10.1007/978-0-387-73003-5_293)
- [185] Marasinou, C., Li, B., Paige, J. et al. Improving the Quantitative Analysis of Breast Microcalcifications: A Multiscale Approach. *J Digit Imaging* 36, 1016–1028 (2023). <https://doi.org/10.1007/s10278-022-00751-3>
- [186] Grignaffini, F., Simeoni, P., Alisi, A., & Frezza, F. (2024). Computer-Aided Diagnosis Systems for Automatic Malaria Parasite Detection and Classification: A Systematic Review. *Electronics*, 13(16), 3174. <https://doi.org/10.3390/electronics13163174>