UNIVERSITAT DE BARCELONA

Treball de Fi de Grau

GRAU D'ENGINYERIA INFORMÀTICA

Facultat de Matemàtiques i Informàtica
Universitat de Barcelona

RADIOMICS-BASED ANALYSIS OF CONTRAST-ENHANCED AND DIGITAL MAMMOGRAPHY FOR BREAST CANCER CLASSIFICATION

Júlia Calderón Estébanez

Director: Dr. Oliver Díaz Montesdeoca
Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, 10 de juny de 2025

Abstract

Contrast-Enhanced Spectral Mammography (CESM) is an advanced imaging modality that enhances breast cancer detection by combining conventional digital mammography (DM) with CESM obtained through intravenous contrast administration. This dual approach provides both morphological and functional information, improving lesion visibility, particularly in patients with dense breast tissue. Radiomics, a rapidly evolving field in medical imaging, allows the extraction of high-dimensional quantitative features from medical images, capturing information about tumour phenotype, texture, and heterogeneity that may not be visually apparent.

This thesis investigates the application of radiomics analysis to CESM images with the aim of improving breast cancer classification. A key focus is on comparing the diagnostic performance of radiomics features derived separately from CESM and DM images, as well as evaluating the added value of combining both sets of features. Radiomics features are extracted and analysed using statistical and traditional machine learning techniques to assess their effectiveness in distinguishing between benign and malignant lesions. Furthermore, the study explores the development of predictive models based on these features and identifies the most relevant biomarkers for tumour classification.

By systematically evaluating radiomics features derived from CESM, DM, and their combination, this research aims to determine the most effective imaging strategy for accurate breast cancer classification. The findings may support more informed clinical decision-making and contribute to the advancement of personalized diagnostic approaches in breast cancer care.

Results show that models based on CESM images consistently outperformed those based on DM and combined data, confirming the hypothesis that contrast-enhanced imaging yields more informative radiomic features for tumour classification.

Abstract

La Mamografía Espectral con Contraste (CESM) es una modalidad de imagen avanzada que mejora la detección del cáncer de mama al combinar la mamografía digital convencional (DM) con CESM, obtenidas mediante la administración intravenosa de contraste. Este enfoque dual proporciona información tanto morfológica como funcional, mejorando la visibilidad de las lesiones, especialmente en pacientes con tejido mamario denso. La radiómica, un campo en rápida evolución dentro de la imagen médica, permite extraer características cuantitativas de alta dimensión de las imágenes, capturando información sobre el fenotipo tumoral, la textura y la heterogeneidad que puede no ser visible a simple vista.

Esta tesis investiga la aplicación del análisis radiómico a la CESM con el objetivo de mejorar la detección y clasificación del cáncer de mama. El estudio se centra en comparar el rendimiento diagnóstico de las características radiómicas derivadas por separado de las imágenes CESM y DM, así como en evaluar el valor añadido de combinar ambas. Se extraen y analizan características radiómicas mediante técnicas estadísticas y modelos de aprendizaje automático para valorar su efectividad en la diferenciación entre lesiones benignas y malignas. Además, se desarrollan modelos predictivos y se identifican los biomarcadores más relevantes para la clasificación tumoral.

Mediante una evaluación sistemática de las imágenes CESM, DM y su combinación a través de radiómica y aprendizaje automático, esta investigación busca determinar la estrategia de imagen más eficaz para una clasificación precisa del cáncer de mama. Los resultados podrían respaldar una toma de decisiones clínicas más informada y contribuir al avance de enfoques diagnósticos personalizados en el cuidado del cáncer de mama.

Al final de la tesis, observamos que resultados muestran que los modelos basados en imágenes CESM superaron de forma consistente a aquellos basados en imágenes DM y datos combinados, lo que confirma la hipótesis de que las imágenes con contraste aportan características radiómicas más informativas para la clasificación tumoral.

Resum

La Mamografia Espectral amb Contrast (CESM) és una modalitat d'imatge avançada que millora la detecció del càncer de mama mitjançant la combinació de la mamografia digital convencional (DM) amb CESM, obtingudes mitjançant l'administració intravenosa de contrast. Aquest enfocament dual ofereix informació morfològica i funcional, millorant la visibilitat de les lesions, especialment en pacients amb teixit mamari dens. La radiómica, un camp emergent en la imatge mèdica, permet extreure característiques quantitatives d'alta dimensió que capturen informació sobre el fenotip tumoral, la textura i l'heterogeneïtat, sovint no visible a simple vista.

Aquesta tesi investiga l'aplicació de l'anàlisi radiómic a la CESM amb l'objectiu de millorar la detecció i classificació del càncer de mama. L'estudi es centra a comparar el rendiment diagnòstic de les característiques radiómiques obtingudes per separat de les imatges CESM i DM, i a avaluar el valor afegit de la seva combinació. S'extreuen i analitzen característiques radiómiques mitjançant tècniques estadístiques i models d'aprenentatge automàtic per determinar la seva efectivitat en la diferenciació entre lesions benignes i malignes. A més, es desenvolupen models predictius i s'identifiquen els biomarcadors més rellevants per a la classificació tumoral.

Mitjançant una avaluació sistemàtica de les imatges CESM, DM i la seva combinació amb radiómica i aprenentatge automàtic, aquesta recerca pretén determinar l'estratègia d'imatge més eficaç per a una classificació precisa del càncer de mama. Els resultats podrien contribuir a una presa de decisions clíniques més informada i a l'avanç d'enfocaments diagnòstics personalitzats en l'atenció al càncer de mama.

Al final de la tesi observem que els resultats mostren que els models basats en imatges CESM van superar de manera consistent els basats en imatges DM i en dades combinades, confirmant la hipòtesi que la imatge amb contrast proporciona característiques radiòmiques més informatives per a la classificació tumoral.

Acknowledgements

I would like to sincerely thank Dr. Oliver Díaz Montesdeoca, for his guidance, support, and constructive feedback throughout the development of this project. His expertise and patience were fundamental to the completion of this work.

I am also grateful to my friends for their support, encouragement, and thoughtful discussions during these years. Their presence made this experience more enjoyable.

Finally, I want to express my deepest gratitude to my family for their unwavering support, patience, and belief in me throughout this journey. Their encouragement has been a constant source of strength.

Table of Contents

Table of Figures

Chapter 1

1. Introduction

1.1 Motivation

Cancer is one of the most pressing public health challenges of our time, affecting millions of individuals worldwide and placing a substantial burden on healthcare systems, economies, and societies. It is a leading cause of death globally, accounting for approximately one in six deaths, and its incidence continues to rise due to factors such as population ageing, lifestyle changes, and environmental exposures, (World Health Organization, 2025). The complexity of cancer lies not only in its biological diversity but also in the social and economic disparities that affect prevention, diagnosis, and treatment outcomes.

Among all cancer types, breast cancer is the most diagnosed one (Figure 1 left), particularly among women, with over 2.3 million new cases reported globally in 2020, (World Health Organization, 2025). However, as we can see on Figure 1 right, when considering mortality, lung cancer remains the leading cause of cancer-related deaths, followed by colorectal, liver, and stomach cancers. These statistics highlight the critical need for effective cancer control and prevention strategies across all stages of the disease, from early detection and accurate diagnosis to personalized treatment and survivorship care.

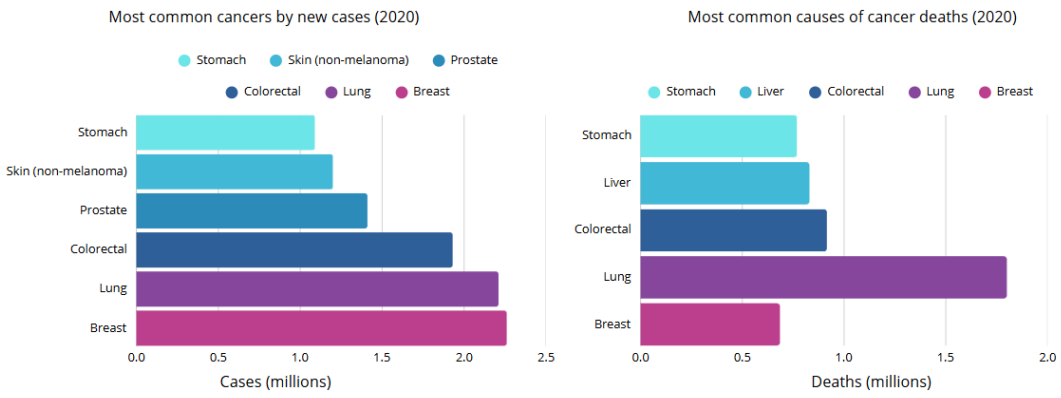

Figure 1: (Right) Most common types of cancer and cancer lethality by cancer type for both sexes worldwide. (Left) Incidence of cancer while left figure represents mortality. (World Health Organization, 2025).

Recognizing the urgency of coordinated action, the European Commission has established the EU Mission on Cancer, one of five flagship research and innovation missions under the Horizon Europe framework (Figure 2). This initiative aims to save more than 3 million lives by 2030 through improved prevention, diagnostics, treatment, and quality of life for cancer patients and survivors. The mission promotes multidisciplinary research and the translation of scientific advances into policy and clinical practice, with a strong emphasis on data-driven innovation and equitable access to care across Europe.

Figure 2: EU Missions: Adaptation to climate change, Cancer, restore our Ocean and waters, climate-neutral and smart cities, a soil deal for Europe. (European Commission, 2021).

The growing importance of cancer research is reflected in the increasing investment in novel diagnostic tools, including advanced imaging technologies and artificial intelligence. These innovations are critical to enabling earlier and more accurate detection, better characterization of tumours, and the development of personalized treatment plans. Within this broader context, breast cancer, due to its high incidence and growing survival rates, represents a key area where technological advances can significantly impact patient outcomes.

1.2 Objectives

The **primary goal** of this thesis is to investigate the application of radiomics analysis in contrast-enhanced spectral mammography (CESM) for improved detection and classification of breast cancer. In addition, the study aims to compare the diagnostic performance of CESM, digital mammography (DM), and a combination of both feature sets, in order to identify the most effective imaging modality or combination for tumour classification. The central hypothesis is that radiomic features extracted from CESM images will provide more informative and discriminative patterns for breast cancer classification than those derived from DM alone. While the integration of CESM and DM features is also evaluated, its potential benefit remains uncertain and is therefore approached as an exploratory component of the analysis.

The **specific objectives** include:

1. To extract and analyse radiomics features from CESM images, digital mammography (DM) images, and both contrast-enhanced and DM components.

2. To evaluate the effectiveness of radiomics features derived separately from CESM and DM images in differentiating benign from malignant breast lesions.
3. To investigate whether the integration of features from both CESM and DM images leads to improved classification performance compared to using either modality alone.
4. To develop and validate machine learning models using the extracted radiomic features, assessing their predictive accuracy and robustness in breast cancer detection.
5. To identify the most significant radiomics features through statistical analysis and feature selection techniques.

1.3 Project timeline

At the start of this work, several tasks were defined to be able to address all the previously define objectives with the time and resources available, as described below.

State of art:

- Literature review on breast cancer detection techniques, CESM imaging, and radiomics applications. Reviewing existing radiomics-based feature extraction and classification methodologies.
- Objectives and problem definition.
- Dataset acquisition: Obtaining the CESM dataset from the Cancer Imaging Archive and verify data availability, formats and metadata.

Implementation:

Data preprocessing and feature extraction:

- Image Preprocessing: Standardize CESM images (normalization, contrast enhancement if needed) and segment regions of interest (ROIs) using manual/automated techniques.
- Radiomics Feature Extraction: Extracting a comprehensive set of radiomics features (texture, shape and intensity).
- Performing feature selection to reduce dimensionality.

Model implementation:

- Developing classification models: Selecting the machine learning models that adapts the best and training the models using the extracted radiomics features...

Evaluation and interpretation:

- Comparing the results against traditional CESM-based breast cancer detection.

Writing and corrections:

- Documenting all the results and discussions.
- Revise the thesis based on feedback from the tutor.

Submission

The following Gantt chart (Figure 3) visually illustrates the project's timeline. It serves as a powerful tool for planification and monitoring the progress of the project, ensuring it reaches its intended stages by specified deadlines.

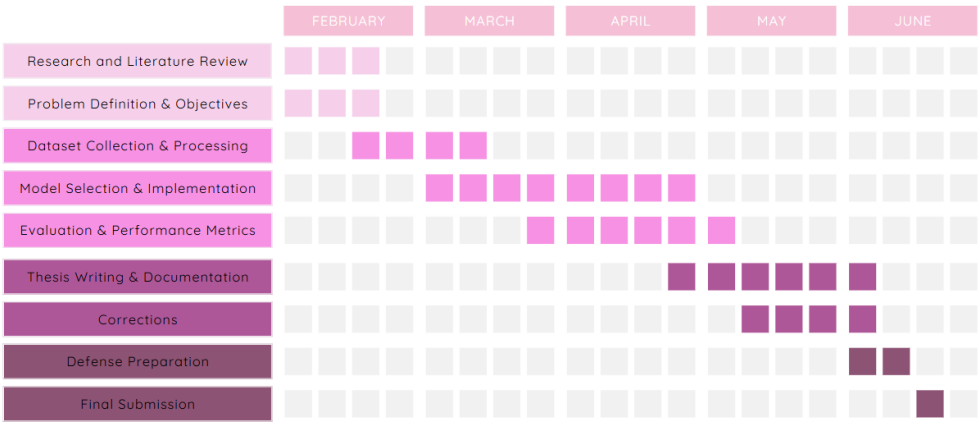

Figure 3: Gantt Chart.

Chapter 2

2. Background and literature review

2.1 Breast cancer

As we have seen on Figure 1, breast cancer is one of the most prevalent and life-threatening diseases among specially women worldwide, affecting millions of individuals each year. Early detection is crucial for improving survival rates, as early diagnosis supports more effective treatment before the disease progresses to advanced stages.

Breast cancer occurs when abnormal cells in the breast grow uncontrollably, forming a tumour that can be benign or malignant. The latter affecting to normal tissue and affecting its normal functions.

As depicted in Figure 4, the breast is an organ composed of lobules, which produce milk, ducts that transport the milk to the nipple, and surrounding fatty and connective tissue. Cancer most commonly originates in the ducts (ductal carcinoma) or lobules (lobular carcinoma) and can either remain localized or spread to other parts of the body through the lymphatic system and bloodstream.

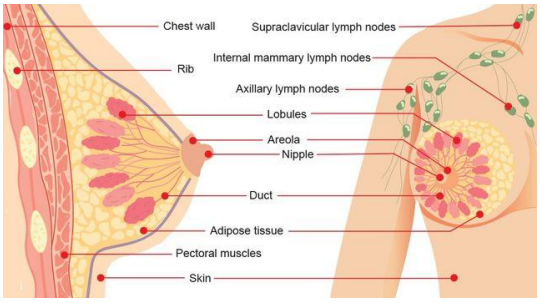

Figure 4: Breast anatomy.
(Dr. Mary Ling, n.d.).

The exact cause of breast cancer remains unknown, but numerous risk factors contribute to its development. Genetic mutations play a significant role, with BRCA1 and BRCA2 gene mutations being the most well-known hereditary factors associated with increased susceptibility. A family history of breast cancer further raises the likelihood of developing the disease. Beyond genetics, hormonal and reproductive factors also influence risk, such as early onset of menstruation, late menopause, hormone replacement therapy, and delayed pregnancy. Additionally, lifestyle choices like sedentary lifestyle, alcohol

consumption, smoking, and large exposure to radiation have been linked to higher breast cancer incidence. *(World Health Organization, 2024) (Borghild Løyland, 2024)*.

Breast cancer is classified into several types based on its origin and molecular characteristics. Ductal carcinoma in situ (DCIS) is a non-invasive form, while invasive ductal carcinoma (IDC) is the most common type that spreads into surrounding tissues. Lobular carcinoma in situ (LCIS) increases the risk of developing invasive cancer, whereas invasive lobular carcinoma (ILC) originates in the milk-producing lobules and can metastasize. On a molecular level, breast cancer is further categorized based on the presence or absence of hormone receptors and HER2 protein expression. Hormone receptor-positive cancers (ER+/PR+) grow in response to oestrogen or progesterone and can be treated with hormone therapies. HER2-positive breast cancer is an aggressive form but responds well to targeted treatments like trastuzumab. Triple-negative breast cancer (TNBC) lacks hormone receptors and HER2 expression, making it difficult to treat with conventional hormone or targeted therapies, often requiring chemotherapy.

The symptoms of breast cancer vary but commonly include a new lump or thickening in the breast or underarm, changes in breast shape or size, skin dimpling, nipple retraction, and abnormal discharge. Some cases remain asymptomatic, highlighting the importance of routine population screening. Early detection is critical, and the disease is staged using the TNM system, which evaluates tumour size (T), lymph node involvement (N), and metastasis (M). Stage 0 refers to non-invasive cancer, while Stage I to III indicate progressive local or regional spread. Stage IV, also known as metastatic breast cancer, signifies that the cancer has spread to distant organs such as the lungs, liver, or bones. Both mortality and incidence rates for each stage is presented in Figure 5.

Treatment options depend on the stage and subtype of breast cancer. Surgery is often the first step, with lumpectomy being a breast-conserving option and mastectomy involving the removal of one or both breasts. Radiation therapy follows surgery in many cases to destroy any remaining cancer cells. Systemic treatments include chemotherapy, hormone therapy, targeted therapy, and immunotherapy, tailored to the cancer's molecular profile. Advances in medical research have significantly improved survival rates, especially for localized breast cancer, where five-year survival exceeds 99%. However, metastatic breast cancer remains challenging, with a five-year survival rate of around 30%.

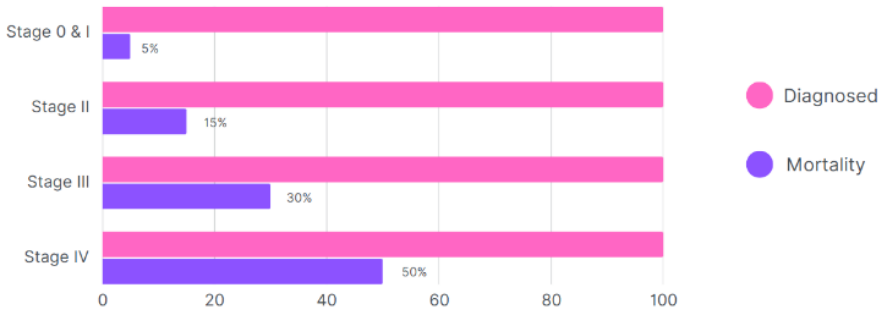

Figure 5: Mortality rate for 5 years or more after diagnosis on different breast cancer stages. (Cancer Research UK, 2021).

2.2 Traditional breast cancer detection methods

Early and accurate detection of breast cancer is critical for improving survival rates and treatment outcomes. Traditional diagnostic methods have been developed over decades, allowing physicians to identify and classify breast cancer based on imaging, clinical examination, and histopathological analysis. These methods serve as the foundation for modern imaging techniques, including the advanced field of Radiomics Analysis of CESM or dynamic contrast enhanced magnetic resonance imaging (DCE-MRI).

2.2.1 Clinical Breast Examination

A clinical breast examination (CBE) is a physical inspection performed by a healthcare provider to detect lumps, abnormalities, or other signs of breast cancer. This involves palpation of the breast tissue and nearby lymph nodes to look for changes in texture, size, or mobility of any masses. While CBE is a simple and cost-effective method, it is subjective and lacks the sensitivity to detect small or deep-seated tumours, especially in dense breast tissue.

2.2.2 Breast Self-Examination

Breast self-examination (BSE), shown on Figure 6, is a method where individuals check their own breasts for lumps, changes in size or shape, or nipple abnormalities. While BSE promotes self-awareness and early symptom recognition, it is not a reliable standalone screening tool. Many breast cancers detected through BSE are already at a more advanced stage, which is why imaging techniques are preferred for early diagnosis.

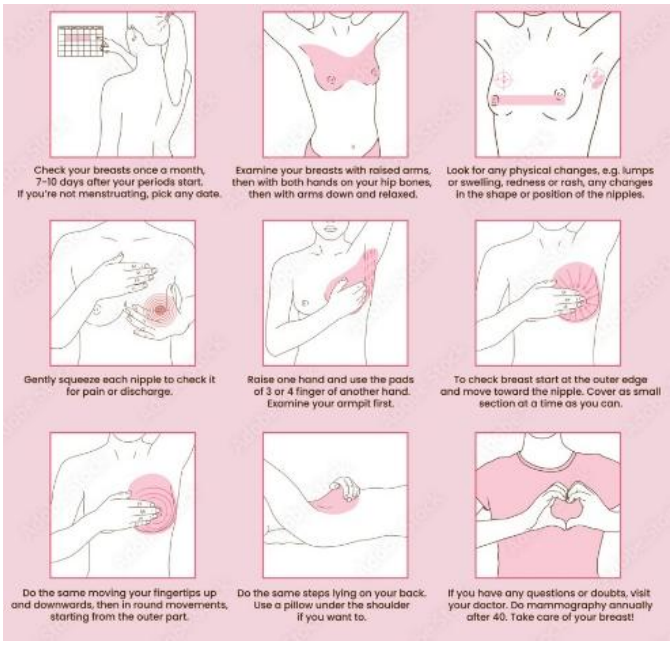

Figure 6: Breast self-examination.
(Mona Moon Naturals, 2023).

2.2.3 Mammography

Mammography is the gold standard for breast cancer screening and involves low-dose X-ray imaging of the breast (Figure 7). It helps detect tumours before they are palpable and identifies microcalcifications, which can be an important indicator of early-stage cancer. There are two main types of mammography:

- Screening Mammography: Used for routine examination in asymptomatic women, typically recommended annually or biennially for women over 40 or those at high risk.

- Diagnostic Mammography: Conducted when abnormalities are detected in a screening mammogram or during a clinical examination. This involves additional imaging views to provide a more detailed assessment.

The effectiveness of mammography is highest in women over 50. However, in younger women with dense breast tissue, it may be less sensitive. In Spain, organized breast cancer screening typically begins at age 50 and continues until age 69, although some regions may start invitations at 45 or 40.

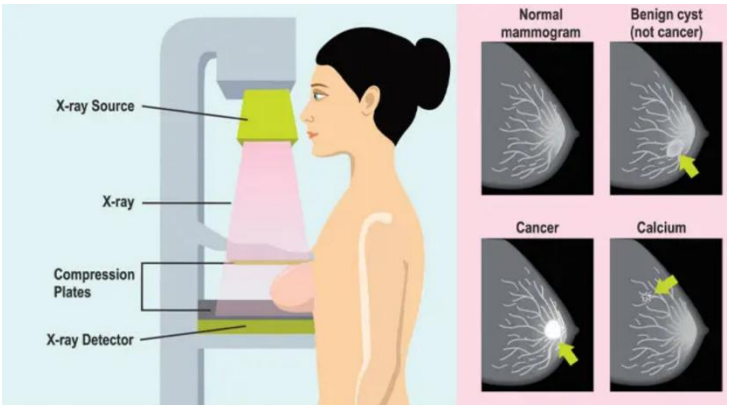

Figure 7: Mammography procedure and possible findings.
(Dr. Chan Ching Wan, 2021).

2.2.4 Ultrasound (Sonography)

Breast ultrasound uses high-frequency sound waves to create images of breast tissue. It is commonly used to differentiate between solid tumours and fluid-filled cysts, evaluate breast abnormalities found on mammography or during a physical exam, and guide biopsies by providing real-time imaging for precise needle placement. (Figure 8).

Ultrasound is particularly useful for younger women with dense breast tissue, since mammography could miss some abnormalities in these cases. However, it is less effective in detecting microcalcifications, which can be an early sign of cancer.

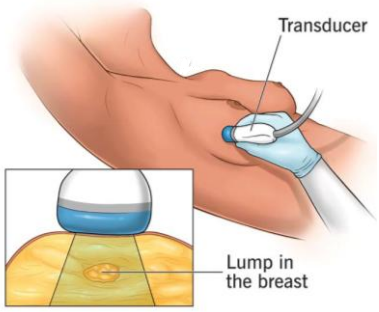

Figure 8: Breast ultrasound procedure.
(Cleveland Clinic, 2024).

2.2.5 Magnetic Resonance Imaging

Breast MRI is a highly sensitive imaging tool that provides detailed images using magnetic fields and contrast agents. It is recommended in specific cases, including:

- High-risk patients (e.g., BRCA1/BRCA2 mutation carriers).
- Evaluation of tumour extent in patients already diagnosed with breast cancer.

- Post-surgical assessment to check for residual cancer.
- Assessment of breast implants for rupture or abnormalities.

Although MRI has high sensitivity, it also has a high false-positive rate, leading to unnecessary biopsies. *(American Family Physician, 2018)*. Additionally, MRI is expensive and not widely available in all healthcare settings.

2.2.6 Biopsy

A biopsy is the only method that can provide a definitive breast cancer diagnosis. It involves extracting a small tissue sample from the patient and send it for laboratory analysis to determine whether the tissue is malignant. There are three different types of breast biopsies (Figure 9):

- Fine Needle Aspiration (FNA): A thin needle is used to extract fluid or small tissue samples. This is a quick and minimally invasive technique but may not provide enough tissue for a definitive diagnosis.
- Core Needle Biopsy (CNB): A larger needle extracts cylindrical tissue samples, allowing for more detailed pathological evaluation.
- Surgical Biopsy: When a larger portion or the entire abnormal area needs to be removed for analysis.

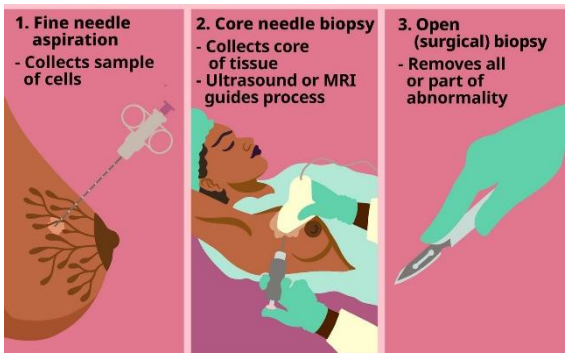

Figure 9: Types of breast biopsies.
(Corey Whelan , 2025).

2.2.7 Molecular and Genetic Testing

Molecular and genetic testing have become important tools in breast cancer diagnostics. These tests help determine an individual's genetic predisposition and tumour characteristics. Common tests include:

- BRCA1/BRCA2 genetic testing to determine hereditary breast cancer risk.

- Hormone receptor testing (ER/PR status) to guide hormone therapy decisions.
- HER2 testing to determine eligibility for targeted treatments like trastuzumab.
- Gene expression profiling (e.g., Oncotype DX, MammaPrint) to predict the likelihood of cancer recurrence and guide chemotherapy decisions.

2.2.8 Limitations of traditional methods

As we have seen, traditional breast cancer detection methods, while effective, have several limitations that can impact their accuracy, accessibility, and overall reliability. Mammography, one of the most widely used screening tools, has reduced sensitivity in women with dense breast tissue. Since dense tissue appears white on a mammogram, similar to tumours, distinguishing between normal and abnormal growths becomes challenging, increasing the risk of missed diagnoses.

Ultrasound and MRI are commonly used as supplementary imaging techniques, however, they also present some disadvantages. While highly sensitive, they often produce false positives, detecting abnormalities that turn out to be benign, resulting in unnecessary biopsies and anxiety for patients. MRI, despite being an effective tool for high-risk individuals, is expensive and not widely accessible, with insurance coverage varying across different healthcare systems. Ultrasound, on the other hand, is highly operator-dependent, meaning its accuracy can fluctuate based on the skill and experience of the technician or radiologist.

Biopsies considered the gold standard for confirming a breast cancer diagnosis, also present certain limitations. Since it is an invasive procedure, can cause discomfort, pain, and in some cases, complications such as bleeding, infection, or scarring. Furthermore, biopsy samples do not always provide conclusive results, sometimes requiring repeat procedures or additional testing to confirm a diagnosis.

Molecular and genetic testing, while offering valuable insights into an individual's cancer risk and tumour characteristics, come with financial and accessibility challenges. These tests, are expensive and not always covered by insurance, making them inaccessible to many patients. Additionally, genetic testing may not be necessary or beneficial for all individuals, as its relevance depends on family history and other risk factors. Results from these tests can also be complex, sometimes leading to uncertainty in treatment decisions.

These limitations highlight the need for continuous advancements in breast cancer screening and diagnosis. Emerging technologies such as artificial intelligence-assisted imaging, liquid biopsies, and personalized risk assessment tools hold promise in improving early detection, reducing unnecessary interventions, and making breast cancer screening more accurate and accessible to everyone.

2.3 Other detection methods

As imaging technology has advanced, several additional modalities have emerged to complement traditional breast cancer detection methods such as 2D mammography, MRI or ultrasound. These newer techniques offer improved visualization, functional insights, and better assessment of disease spread, particularly in complex or ambiguous cases.
While not all these methods are used for routine screening, they provide valuable information in specific clinical context, such as in women with dense breast tissue, for staging cancer, or for monitoring metastatic disease.

2.3.1 Digital Breast Tomosynthesis (DBT)

Also known as 3D mammography, is an advanced imaging technique used primarily for diagnosis and breast cancer screening in some countries. Unlike traditional 2D mammography that takes a single planar image, DBT captures multiple low-dose X-ray images of the breast from limited angles (Figure 10). These image projections are then reconstructed into thin, layered slices, giving a pseudo three-dimensional view of the breast tissue (i.e., a stack of 2D images representing different depths of the breast). This makes it easier to detect abnormalities that might be hidden in overlapping tissue, particularly in women with dense breasts. DBT has been shown to improve cancer detection rates and reduce the number of false positives, meaning fewer women are called back for unnecessary additional testing. While it does involve slightly more radiation than standard mammography, it remains within safe exposure levels.

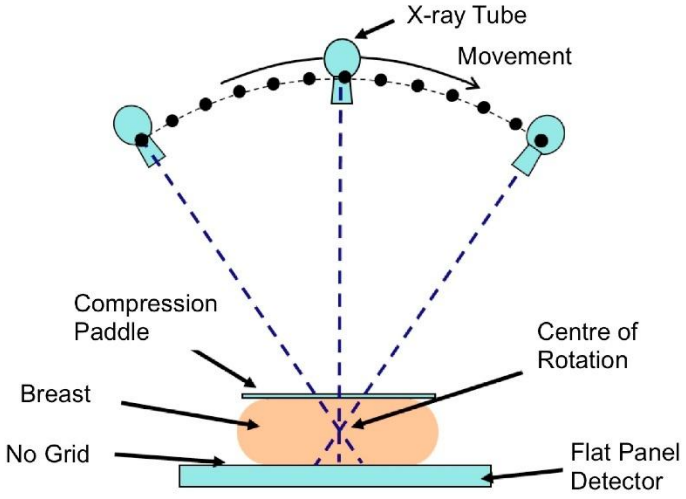

Figure 10: Digital Breast Tomosynthesis diagram procedure. (Royal Surrey County Hospital, 2018).

2.3.2 Positron Emission Tomography (PET)

PET is a functional imaging technique that provides information about the metabolic activity of tissues. In breast cancer evaluation (Figure 11), the most used PET tracer is fluorodeoxyglucose (FDG), a radioactive form of glucose. Since cancer cells tend to grow

and divide rapidly, they consume more glucose than normal cells. After FDG is injected into the bloodstream, it accumulates in areas with high metabolic activity, such as tumours, which can then be detected by the PET scanner. PET imaging is especially valuable for staging breast cancer, detecting distant metastases, and monitoring treatment response. It offers a whole-body view, which is useful for identifying cancer spread to organs like the lungs, liver, or bones. However, PET does not provide detailed anatomical images and is therefore often combined with a CT scan (PET/CT) to correlate metabolic activity with precise anatomical locations.

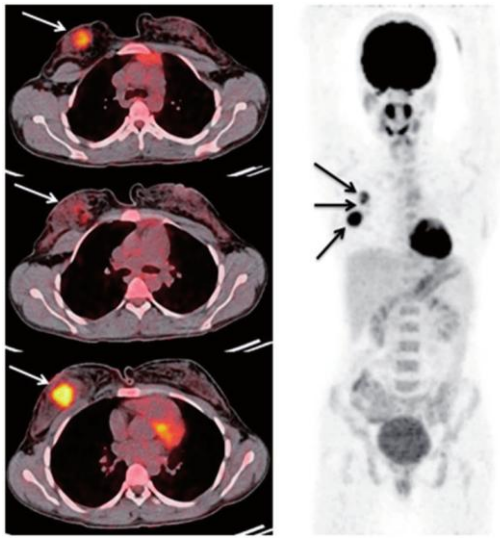

Figure 11: Positron Emission Tomography (PET) results example.
(Ergul, Nurhan & Kadioglu, Huseyin & Yildiz, Seyma & Yucel, Serap & Gucin, Zuhal & Erdogan, Ezgi & Aydin, Mehmet & Muslumanoglu, Mahmut., 2014).

2.3.3 Single Photon Emission Computed Tomography (SPECT)

SPECT is another nuclear medicine imaging technique, but it uses different types of radioactive tracers and a gamma camera to capture 3D images (Figure 12). In breast cancer, SPECT can be used with tracers like technetium-99m to evaluate bone metastases or to visualize sentinel lymph nodes prior to surgery. SPECT is particularly useful in evaluating skeletal involvement in advanced breast cancer cases, often in the form of a bone scan. While SPECT also provides functional information, its resolution is generally lower than PET, and it is used less frequently for soft tissue imaging. Like PET, it is not used for routine breast cancer screening but serves as an important tool in staging and follow-up when specific clinical questions arise.

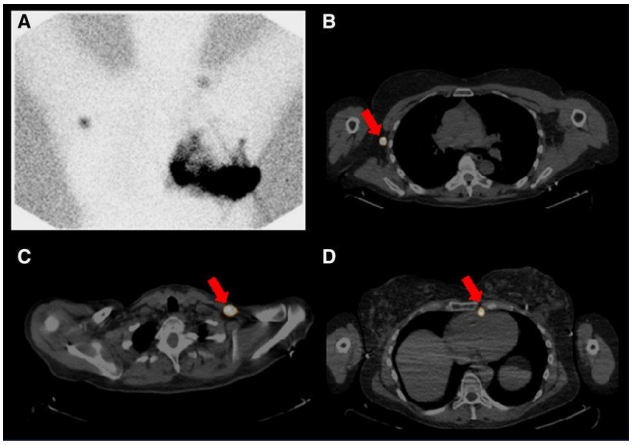

Figure 12: Single Photon Emission Computed Tomography results example.
(Marino, Maria Adele & Avendaño, Daly & Zapata-Julián, Pedro & Riedl, Christopher & Pinker, Katja, 2019).

2.3.4 Computed Tomography Scan (CT)

CT is a type of imaging that uses X-rays to create detailed cross-sectional images of the body (Figure 13). While CT scans are not usually used for the initial detection of breast cancer, they play an important role in evaluating the spread of the disease to other organs, such as the lungs, liver, or bones. CT is fast, widely accessible, and effective for identifying larger tumours or metastases in other parts of the body. However, it is not as sensitive as DBT or MRI for detecting small or early breast cancers, and its use involves a relatively higher dose of radiation. For this reason, CT is generally reserved for specific clinical indications rather than routine screening.

Figure 13: Example of a 2D slide of a CT scan result. Contour of the patient is shown in yellow whereas suspicious breast lesion is depicted in red.
(Janusz Skowronek, 2011).

2.4 Mammographic views

In CESM, different mammographic views are taken to obtain a comprehensive evaluation of the breast tissue from various angles, as we can see in Figure 14. The Craniocaudal (CC) view is performed on both breasts in screening programs, providing a top-down image where the breast is compressed horizontally between two plates. This view captures the

breast tissue from the nipple towards the chest wall, offering a symmetrical image that helps compare both breasts and detect abnormalities in the central and lower regions.

The Mediolateral Oblique (MLO) view is taken at an angle, typically around 45 degrees with respect to CC x-ray tube position, to visualize more of the breast tissue, especially the upper outer quadrant, where most breast cancers are likely to develop. This angled perspective allows for better imaging of the tissue close to the chest wall and the axillary (underarm) area, which is not always visible in the CC view. The MLO view is particularly useful for detecting abnormalities near the lymph nodes or deeper structures.

Both views are essential as they complement each other, ensuring that all areas of the breast are thoroughly examined. The combination of CC and MLO views enhances the likelihood of detecting small lesions or suspicious areas that may not be visible from a single angle. These standardized projections are crucial in breast imaging, providing detailed and consistent information to aid in accurate diagnosis.

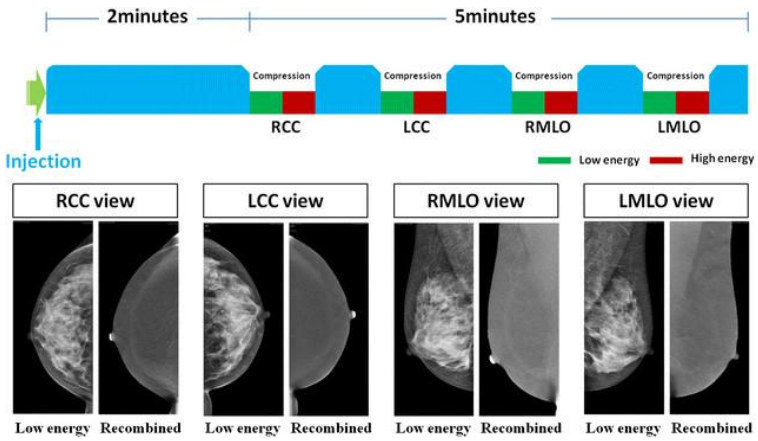

Figure 14: Standard mammographic views used in clinical practice. (Wang, Q., Li, K., 2016).

2.5 BI-RADS

In breast imaging, obtaining high-quality images from different angles is essential for accurate diagnosis, which is why standardized mammographic views such as RCC, LCC, RMLO, and LMLO are used. These views ensure comprehensive visualization of the breast tissue, allowing radiologists to detect abnormalities that may not be visible from a single perspective. However, once these images are obtained, they must be interpreted systematically to determine whether the findings are normal, benign, or suspicious. This is where the Breast Imaging Reporting and Data System (BI-RADS) comes into play, *(American College of Radiology, 2013)*.

BI-RADS provides a structured classification system that helps radiologists categorize breast lesions based on their imaging characteristics, ensuring consistency in diagnosis and guiding appropriate follow-up or treatment. Together, standardized mammographic views and BI-RADS classification form the foundation of breast cancer screening and diagnostic evaluation, enabling early detection and improving patient outcomes.

When a breast imaging study is performed, the findings are categorized into a BI-RADS classification ranging from 0 to 6 (see Table 1). A BI-RADS 0 assessment indicates that the evaluation is incomplete, meaning additional imaging, such as magnified views, ultrasound, or prior imaging for comparison, is required before reaching a final conclusion. A BI-RADS 1 result signifies a completely normal and negative finding, where no abnormalities are detected, and routine screening continues as recommended. Similarly, BI-RADS 2 also represents a benign finding, such as a simple cyst, fibroadenoma, or benign calcifications, which require no additional follow-up beyond regular screening.

A BI-RADS 3 classification suggests a finding that is probably benign, with a very low likelihood of malignancy (less than 2%). However, due to the small possibility of change over time, short-term follow-up imaging, usually in six months, is recommended to monitor for stability. If the finding remains unchanged, routine screening resumes. When an imaging study results in a BI-RADS 4 assessment, it means that the abnormality is suspicious for malignancy, and a biopsy is typically recommended to determine if the lesion is cancerous. The likelihood of malignancy in this category varies, which is why it is further divided into three subcategories: low suspicion (4A), moderate suspicion (4B), and high suspicion (4C).

A BI-RADS 5 assessment indicates that the abnormality is highly suggestive of malignancy, with a greater than 95% probability of being cancerous. Immediate biopsy and further oncological evaluation are necessary. Finally, BI-RADS 6 is used when a patient has already been diagnosed with breast cancer, and imaging is being conducted for treatment planning, such as assessing tumor response to therapy.

Beyond classification, BI-RADS also includes standardized terminology for describing masses, calcifications, asymmetries, and other findings, ensuring that radiologists provide clear, uniform descriptions. Additionally, the system plays a crucial role in risk stratification, guiding whether a patient requires routine screening, short-term monitoring, or immediate biopsy. By eliminating ambiguity in reports, BI-RADS enhances early cancer detection, reduces unnecessary interventions for benign findings, and helps streamline patient care.

Table 1: BI-RADS Categories.

Category		Management	Likelihood of cancer
0	Need additional imaging or prior examinations	Recall for additional imaging and/or await prior examinations	n/a
1	Negative	Routine screening	Essentially 0%
2	Benign	Routine screening	Essentially 0%
3	Probably benign	Short interval-follow-up (6 months) or continued	> 0% but ≤ 2%
4	Suspicious	Tissue diagnosis	4a. Low suspicion for malignancy (> 2% to ≤10%) 4b. Moderate suspicion for malignancy (> 10% to ≤50%) 4c. High suspicion for malignancy (> 50% to <95%)
5	Highly suggestive of malignancy	Tissue diagnosis	≥ 95%
6	Rhown biopsy-proven	Surgical excision when clinical appropriate	n/a

2.6 Breast density

The breast is manly composed of adipose (i.e., fat) and glandular tissue. Breast density refers to the proportion of fibroglandular tissue, which includes milk glands and ducts, to fatty tissue. It plays a crucial role in breast imaging as it affects how mammograms appear and influences the ability to detect abnormalities. Breast density varies among individuals and can change over time due to factors such as age, hormonal fluctuations, and genetic predisposition.

Based on the BI-RADS classification, breast density is divided into four categories (i.e., A to D) ranging from almost entirely fatty (i.e., BI-RADS A) to extremely dense (i.e., BI-RADS D). Breasts that contain mostly adipose tissue shows a higher detection rate since fat appears

dark on a mammogram while tumors and other dense structures appear white. Scattered fibroglandular density means that most of the breast is composed of fat, but there are areas of denser tissue that could obscure small abnormalities. When the breast is classified as heterogeneously dense, a significant amount of fibroglandular tissue is present, which can make it challenging to identify malignancies. Extremely dense breasts, on the other hand, contain very little fatty tissue, making it significantly more difficult to detect cancer using traditional mammography.

In conventional mammography, dense breast tissue can complicate cancer detection because both fibroglandular tissue and malignant tumors appear white on an X-ray, creating a masking effect. This issue arises because the attenuation coefficients of tumor and glandular tissue are similar, making it difficult to distinguish between them. This similarity in appearance can result in false negatives, where cancers go undetected, and false positives, where benign dense tissue is mistaken for a suspicious lesion, leading to unnecessary biopsies and patient anxiety. The reduced sensitivity of mammography in dense breasts means that small tumors may be missed, delaying diagnosis and treatment.

2.7 Contrast-Enhanced Spectral Mammography (CESM)

Traditional diagnostic methods, such as mammography and ultrasound, have limitations, particularly when used to detect tumours in women with dense breast tissue. These limitations highlight the need for advanced imaging techniques to improve detection accuracy and reduce human error.

CESM is an innovative and advanced breast imaging modality designed to significantly improve the detection and characterization of breast cancer. It combines traditional mammography with the use of contrast agents to enhance the visibility of tumours. This technique is particularly beneficial in cases where conventional mammography may not be as effective, such as with dense breast tissue or when subtle lesions may be difficult to detect using standard imaging methods. CESM offers a higher degree of sensitivity, making it an increasingly valuable tool in early breast cancer diagnosis.

CESM is typically used in settings where there is a suspicion of breast cancer or when previous imaging results have been inconclusive. It also plays a critical role in assessing tumour biology and behaviour, which is vital for accurate treatment planning and improving patient outcomes.

The innovative aspect of CESM is that it may be easier to spot cancer, because there is better contrast between suspicious areas, with increased blood flow, and normal breast tissue. For people who might have otherwise had an MRI scan, imaging with these technologies will be quicker and may be completed at the breast clinic, instead of in the radiology department. This could release time and resources for MRI in other areas.

2.7.1 How CESM Works

CESM uses iodine-based contrast agents, which are intravenously administered to the patient before the imaging procedure. The iodine in the contrast agent is selectively absorbed by abnormal, cancerous tissue, which is often characterized by increased blood supply due to the process of angiogenesis[1]. This increased blood flow is an indicator of malignant tumours, as they require more nutrients and oxygen to support their rapid growth.

The CESM imaging process begins with the injection of iodine-based contrast agents into the patient's vein. Once administered, the contrast agent circulates through the bloodstream and is absorbed by areas with abnormal blood flow, such as cancerous tissues. After the contrast has been absorbed, CESM uses dual-energy X-ray imaging. The first set of images, captured with low energy, visualizes typical breast structures similar to standard mammography, focusing on dense tissues and revealing anatomical features such as glandular tissue and fatty areas.

Next, a second set of high-energy images is taken to highlight areas where the contrast agent has been absorbed. Cancerous tissues, which absorb more contrast due to increased blood supply, show up as high-intensity areas on the scan. These high-energy images help to pinpoint areas that may indicate malignancy. A subtraction technique is then applied to remove the low-energy images, leaving only the enhanced areas from the high-energy scan. This process highlights contrast-enhanced areas where abnormalities are most likely to be present, resulting in a clear and focused image that makes tumours and lesions more visible. (Figure 15).

By combining the advantages of traditional mammography and the contrast-enhanced imaging, CESM helps improve the sensitivity of breast cancer detection, particularly in complex cases where conventional methods may not provide a definitive diagnosis.

Figure 15: A typical contrast-enhanced spectral mammography (CESM) examination.
a. Low-energy b. High-energy c. Recombined (subtracted).

[1] Is the physiological process through which new blood vessels form from pre-existing vessels.

(U.C.Lalij, & Jeukens, C.R.L.P.N. & Houben, Ivo & Nelemans, P.J. & Engen, Ruben & Wylick, E. & Beets-Tan, Regina & Wildberger, J.E. & Paulis, Leonie & Lobbes, Marc, 2015).

2.7.2 Procedure overview

The procedure lasts approximately 30 minutes.

Upon arrival, the patient will be asked to complete a questionnaire titled the Pre-Contrast Enhanced Spectral Mammography Examination Checklist. This ensures that they are eligible for the procedure.

Following this, the patient will have a consultation with the mammographer (a trained technologist responsible for performing breast X-rays). During this discussion, the mammographer will review the patient's general health and provide a detailed explanation of the procedure.

An intravenous cannula will then be inserted into a vein, typically on the back of the hand or in the crease of the elbow. The patient may experience a brief scratching sensation when the needle is inserted, but once in place, it should not cause further discomfort. A tourniquet may be applied to the upper arm to facilitate the insertion of the needle.

A contrast medium will be administered through the cannula. As the contrast circulates through the body, the patient may experience a warm sensation (this is completely normal and not a cause for concern).

After a short waiting period, the mammographer will ask the patient to remove their hospital gown and stand in front of the X-ray machine. Each breast will be positioned in the machine and gently but firmly compressed by a clear plate. This compression is essential to minimize radiation exposure and prevent image blurring. The compression is released automatically once the X-ray is taken. Each breast may require multiple images from different angles.

The mammogram itself takes only a few minutes, with each breast being compressed for a few seconds per image. While some patients may find the compression uncomfortable, it is brief.

Once all necessary images have been captured, the patient will be asked to wait in the designated waiting area. The cannula will remain in place until the radiologists confirm that the images are of sufficient quality. Once approved, the cannula will be removed, and the procedure will be complete.

2.8 Radiomics

Radiomics is an emerging field in medical imaging that focuses on the extraction of a large number of quantitative features from standard medical images using advanced computational algorithms. (Philippe Lambin, 2012). These features, often imperceptible to the human eye, can capture information about tumour shape, texture, intensity, and spatial relationships within tissues. By converting images into high-dimensional data, radiomics enables a more detailed characterization of tumours, potentially improving diagnostic, prognostic, and predictive accuracy.

The radiomics workflow typically involves several key steps: image acquisition, segmentation of the region of interest (ROI), feature extraction, feature selection, and data analysis or modelling (Figure 16). This process allows for the development of imaging biomarkers that may correlate with underlying pathophysiological processes, tumour heterogeneity, or treatment response. Feature extraction is generally performed using open-source libraries such as PyRadiomics, which adhere to standardized definitions provided by the Image Biomarker Standardization Initiative (IBSI). Radiomic features are commonly grouped into categories such as first-order statistics (describing intensity distribution), shape features (describing geometry), and texture features derived from matrices such as the gray level co-occurrence matrix (GLCM), run length matrix (GLRLM), size zone matrix (GLSZM), and others. Additionally, features can be computed from filtered images, such as those processed with Laplacian of Gaussian (LoG) or wavelet transformations, enabling multiscale representation of tissue heterogeneity.

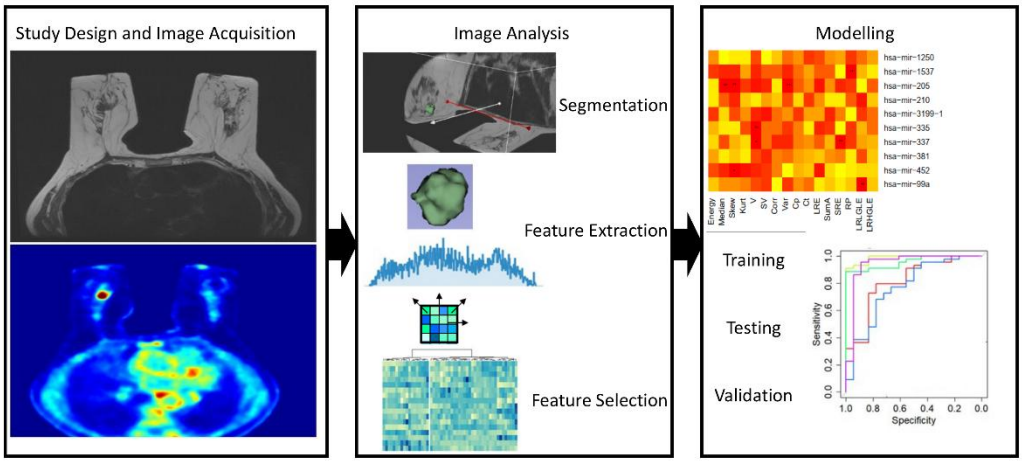

Figure 16: Example of radiomics workflow.
(Francesca Gallivanone,Gloria Bertoli and Danilo Porro., 2022).

In the context of breast cancer, radiomics offers a non-invasive means to enhance the detection and characterization of lesions. When applied to modalities such as CESM, MRI or DM, radiomics can uncover subtle differences in tissue characteristics that may not be

visually apparent. This can lead to improved differentiation between benign and malignant lesions, better risk stratification, and potentially, earlier diagnosis.
Importantly, radiomics has the potential to complement radiological assessments and support clinical decision-making by providing objective, reproducible, and quantitative data.

Despite its promise, radiomics also faces several technical and methodological challenges. Feature reproducibility can be affected by variations in image acquisition protocols, segmentation quality, and preprocessing settings. Moreover, the high dimensionality of radiomic features, relative to typical sample sizes in medical datasets, necessitates careful handling to avoid overfitting and ensure generalizability. Nonetheless, radiomics has been successfully applied across various imaging modalities, including CT, MRI, PET, and ultrasound, and clinical domains, ranging from oncology to neurology and cardiology. *(Liang, Y., Xu, H., Lin, J. et al. , 2025) (Rizzo, S., Botta, F., Raimondi, S. et al. , 2018)*.

2.9 Applications of radiomics in breast imaging

Radiomics has found particularly strong relevance in breast imaging, where it has been applied to improve the detection, characterization, and prognostication of breast cancer. Mammography, as a widely used and non-invasive modality, presents an accessible platform for radiomic analysis, especially given its high spatial resolution and availability in screening programs. Studies have demonstrated that radiomic features extracted from mammographic images can capture microstructural differences between benign and malignant lesions that are not always visible in conventional image interpretation.

Recent research has reinforced this potential. For example, a pilot study focused on CESM demonstrated that texture-based radiomic features could effectively differentiate between benign and malignant breast lesions, offering diagnostic insights that go beyond conventional imaging alone, reporting a mean accuracy of 80% when using SVM model *(Losurdo, 2019)*. Another investigation assessed the ability of CESM-based radiomics to identify triple-negative breast cancer (TNBC). The study revealed that machine learning models trained on CESM-derived features could distinguish TNBC from other molecular subtypes with promising AUC of 0.90 (95% CI: 0.85–0.96), sensitivity of 0.97 and specificity of 0.69 for the combined CC and MLO views when in the test set, suggesting potential for more precise phenotyping and treatment planning *(Zhang Yongxia, 2021)*.

Research in this domain has explored radiomics as a complementary tool to traditional diagnostic approaches. Several studies have shown that radiomic models based on mammography can achieve classification performance comparable to or better than BI-RADS assessments made by expert radiologists. Radiomic features have also been used to predict tumor subtypes, such as triple-negative or HER2-positive cancers, which have distinct treatment pathways and prognostic implications. These findings are echoed in a

scoping review, which not only highlighted the diagnostic value of mammography-based radiomics but also emphasized its role in predicting outcomes such as tumor recurrence, patient survival, and lymph node involvement. For instance, the review cites studies where integrating radiomic features with clinical data improved model performance, yielding increases in AUC of up to 11% for predicting histological grade and lymph node metastasis, and 13% for assessing tumor invasiveness *(Siviengphanom, 2021)*.

Beyond binary classification, radiomics has been applied for predicting axillary lymph node involvement, recurrence risk, and treatment response, thereby extending its utility into personalized oncology. Radiomics has also been applied to other breast imaging modalities, such as DCE-MRI and ultrasound, often in combination with mammography. Multimodal radiomics models have shown improved performance, suggesting that different imaging techniques capture complementary aspects of tumor biology. Some studies have further integrated clinical or genomic data with radiomics features, aiming to develop hybrid models that unify radiomic, genomic, and clinical information for enhanced prognostic accuracy.

However, despite its clinical potential, the adoption of radiomics in breast imaging still faces limitations. Many published models are based on retrospective, single-institution datasets and lack external validation. In addition, there is considerable variability in how radiomic pipelines are implemented, which can affect reproducibility and hinder clinical translation *(Yu Ji, 2019) (Nazmul Ahasan Maruf, 2025)*. Standardized methodologies and rigorous multi-institutional studies are necessary to bridge the gap between research and clinical application.

Chapter 3

3. Materials and methods

3.1 CDD-CESM Dataset

This study uses an open access dataset available from The Cancer Imaging Archive repository *(The Cancer Imaging Archive, 2021)*, which includes both CESM and DM images. The dataset also provides ROIs, which highlight areas with potential abnormalities like tumours, and accompanying diagnostic information (Figure 17).

The dataset is a collection of low-energy images with their corresponding subtracted CESM images gathered from the Radiology Department of the National Cancer Institute, Cairo University, Egypt over the period from January 2019 to February 2021. The images are all high resolution with an average of 2355 × 1315 pixels. Institutional review board approval and patient informed consent to carry out and publish data were obtained from 326 female patients aged from 18 to 90 years. *(Khaled, R., Helal, M., Alfarghaly, O. et al., 2021)*.

Usually, each patient has a total of 8 images, 4 images for each breast side consisting of low energy and subtracted CESM images for each CC and MLO view. However, there are 46 patients with only 4 images as they had mastectomy on a breast side, and 87 patients with missing images as some were not available or removed due to quality concerns. *(Khaled, R., Helal, M., Alfarghaly, O. et al., 2021)*.

Two different machines were used for image acquisition: GE Healthcare Senographe DS and Hologic Selenia Dimensions Mammography Systems. The two machines provide similar quality. *(Khaled, R., Helal, M., Alfarghaly, O. et al., 2021)*.

The image dataset consists of 2,006 images from 326 patients, with a total size of 1.49 GB (comprising 1,003 DM images (0.64 GB) and 1,003 CESM images (0.82 GB)). (Figure 19 and Figure 20 show the count of benign and malignant cases across the dataset, and it's distribution based on image type, respectively).

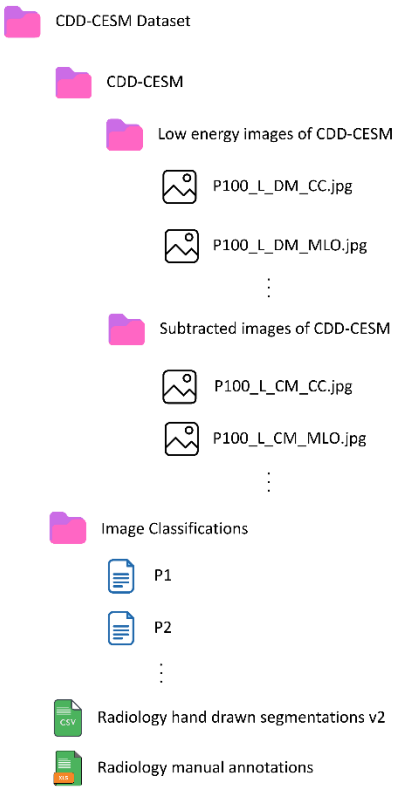

Figure 17: Structure of the folder of the dataset used in the project.

1. Mammography Images on .JPG File format (CESM, DM)

- CESM: These images involve the use of contrast agents, which are injected into the patient to enhance abnormal tissue visibility during the mammogram (subtracted images).
- DM: Digital mammography captures high-resolution images of the breast using electronic sensors instead of film (low energy images).
- File's name structure is (Figure 18):
  - PX: Patient ID.
  - L or R: Side of the breast (Left or Right).
  - CM or DM: Type of imaging (Contrast Mammography or Digital Mammography). (**Note**: Although the dataset refers to CESM images using the abbreviation "CM" in the file names, this should not be confused with Conventional Mammography. Throughout the code and plots, the term "CM" appears due to the dataset's naming convention. However, it is important to emphasize that all such instances are referring to CESM images.)
  - MLO or CC: The specific view taken (either Mediolateral Oblique or Craniocaudal).

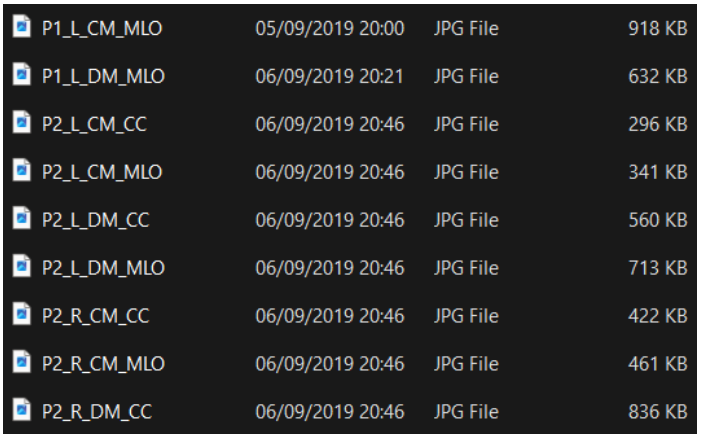

Figure 18: Example entry from the dataset with CESM and DM mammograms merged. The image displays metadata for a mammogram file, including the filename, date of inclusion in the dataset, file type, and image size.

2. Regions of Interest (ROIs)

This CSV file contains annotations for specific areas within the mammogram images that have been identified as ROIs, is the file that contains the different coordinates. These regions typically highlight areas where abnormalities (like tumours, calcifications, or other suspicious features) are present.

3. Clinical Variables

The images are manually-annotated by expert radiologists according to the American College of Radiology Breast Imaging Reporting and Data System (ACR BIRADS) 2013 lexicon for standardized descriptors.

This CSV file contains a variety of metadata about the mammogram images:

Table 2: Clinical variables stored for each mammogram image on a CSV file inside the dataset.

Image name	The identifier of the image (e.g., P1_L_CM_MLO), linking it to the relevant patient and view.
Patient_ID	The unique identifier for the patient undergoing the examination.
Side	Left or right breast.
Type	Type of mammogram (CM or DM).
Age	The patient's age at the time of the mammogram.

Breast density (ACR)	Categorizes the breast tissue density, which impacts how well abnormalities are detected.
BIRADS	The BIRADS score that rates the risk of malignancy, helping to guide clinical decision-making.
Findings	Describes any abnormalities found in the mammogram, such as masses, calcifications, or distortions.
View	The specific view of the mammogram (CC, MLO).
Tags	Any additional categorization or labels related to the mammogram.
Machine	The machine used to perform the mammogram (1 or 2).
Pathology Classification/Follow-up	The final diagnosis based on the mammogram, typically indicating whether the findings are Normal, Benign, or Malignant. It may also include follow-up recommendations, such as further imaging or biopsy.

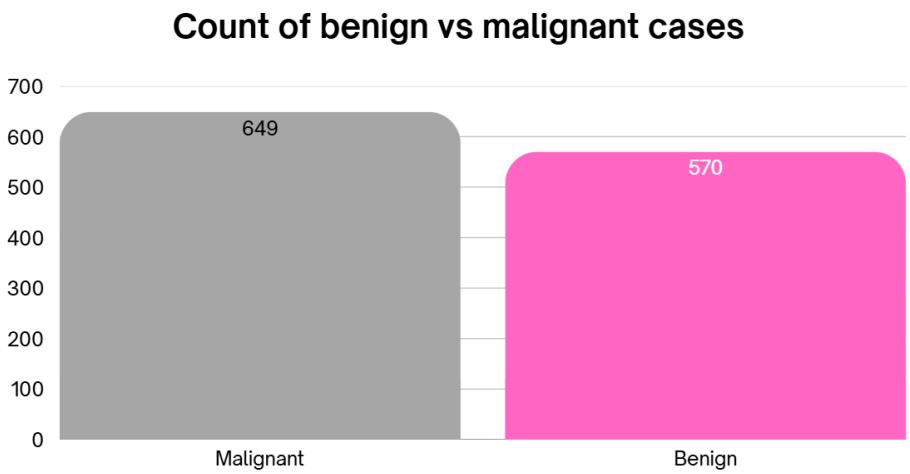

Figure 19: Malignant and benign cases on the dataset.

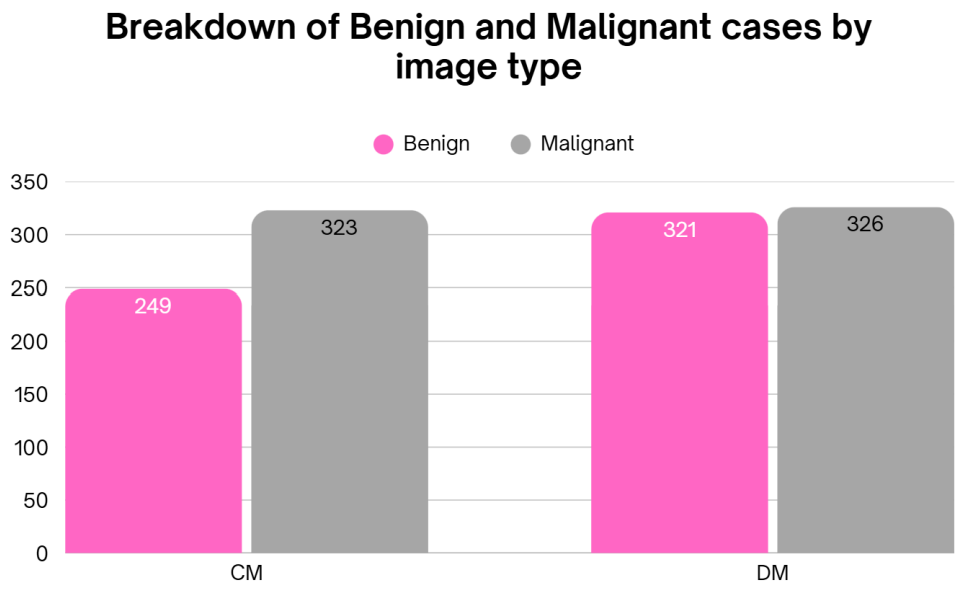

Figure 20: Type of mammography and tumour distribution on the dataset.

3.2 Image preprocessing

Preprocessing plays a crucial role in the analysis of mammograms, particularly in preparing the images for neural networks *(Diaz, 2021)*. Mammograms, like many medical images, often display variations in brightness and contrast, and they can be affected by noise. Neural networks perform best when the input data is consistent and of high quality. Therefore, preprocessing is essential for standardizing the images, ensuring that the neural network can focus on learning the most relevant features without being distracted by inconsistencies or artifacts.

An additional and critical reason for applying preprocessing techniques in this study is the heterogeneity of the dataset, which was compiled from two different mammography machines. These devices may produce images with differing intensity distributions, contrast levels, resolutions, and noise characteristics due to variations in hardware, imaging protocols, or calibration settings. Such discrepancies can significantly affect the consistency of the input data, making it more challenging for the neural network to generalize and learn meaningful patterns. Without adequate preprocessing, these machine-induced differences could introduce bias, reduce model performance, and obscure subtle diagnostic features. By applying standardized preprocessing methods, we aim to harmonize the images across sources, ensuring that the model learns from anatomical and pathological features rather than from scanner-specific artifacts.

Another important aspect of preprocessing is highlighting significant features within the images. In mammograms, subtle differences in breast tissue can indicate abnormalities, and it is vital to make these variations more visible. By enhancing the contrast and adjusting the brightness of the images, preprocessing brings out these subtle details, making it easier for the neural network to detect important patterns that might otherwise go unnoticed.

To achieve these improvements, preprocessing techniques are chosen to maximize their effectiveness. Min-max scaling is one such technique, where the pixel values of an image are rescaled to a defined range, typically between 0 and 1. Neural networks often perform better with input values constrained within a limited range, and this method helps normalize image brightness, preventing variations from confusing the model.

In addition to min-max scaling, z-score normalization was employed. This method standardizes the pixel values by subtracting the mean and dividing by the standard deviation, effectively centring the data around zero. By ensuring that all images share a similar distribution of pixel values, z-score normalization makes the dataset more Gaussian, a property that benefits many types of neural network models.

Another essential technique is CLAHE, or Contrast Limited Adaptive Histogram Equalization. CLAHE specifically enhances the contrast within images, especially in regions

where natural contrast is low. By dividing the image into smaller sections and adjusting the contrast locally, CLAHE uncovers subtle abnormalities that could otherwise remain hidden. This method is particularly well-suited for mammograms, as it enhances image details effectively without excessively amplifying noise, as we can see on Figures 21 and 22 for CESM and DM, respectively. *(Monserrate Intriago-Pazmiño, 2022)*.

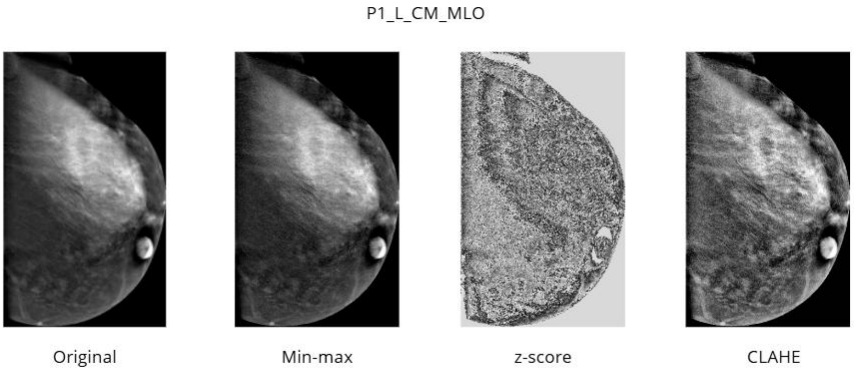

Figure 21: Image preprocessing techniques example for Contrast Mammography.

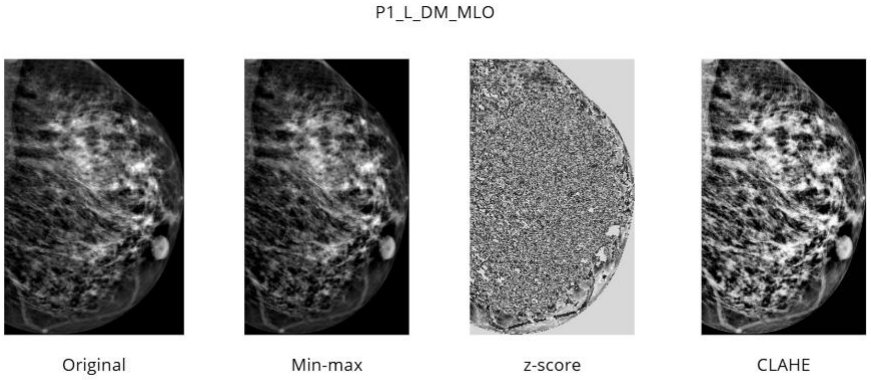

Figure 22: Image preprocessing techniques example for Digital Mammography.

3.3 Radiomics pipeline

Radiomics features were extracted from 2D mammographic pre-processed images using the PyRadiomics 3.1.0 library, based on manually segmented binary masks defining ROIs. Feature extraction followed a standardized parameter configuration, including a fixed bin width (25), 2D enforcement (force2D=True), and the application of both original and filtered image types (Wavelet and Laplacian of Gaussian with σ = 1.0, 3.0, 5.0). Six feature classes were extracted: first-order statistics, gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), gray level dependence matrix (GLDM), and neighboring gray tone difference matrix (NGTDM). The resulting feature matrix, comprising 558 features per image, was saved in CSV format for further analysis.

To evaluate the diagnostic value of these features, a range of supervised machine learning models was implemented in Python using scikit-learn, imbalanced-learn, XGBoost, LightGBM, TensorFlow/Keras, and SHAP. The selected models included Support Vector Machines (SVM), Random Forests, Logistic Regression, K-Nearest Neighbors (KNN), XGBoost, LightGBM, and feedforward neural networks. This selection was made to include both traditional machine learning and deep learning approaches, ensuring a robust comparative analysis across model types. SVMs are well suited for high-dimensional classification tasks, while Random Forests aggregate decision trees to enhance stability and reduce overfitting. Logistic Regression provides a simple, interpretable linear baseline. KNN offers a distance-based approach to classification. XGBoost and LightGBM are advanced gradient boosting methods known for their performance and efficiency on structured data. Feedforward neural networks, built with fully connected layers, batch normalization, and dropout regularization, were used to assess the capabilities of deep learning on radiomics data.

Before training, all features were standardized using z-score normalization, and highly correlated features (Pearson's $r > 0.95$) were removed to reduce redundancy and improve model generalizability. To further mitigate the risk of overfitting and reduce computational cost, a feature resampling strategy was employed during training: at each cross-validation fold, a random subset of radiomic features (around 200 out of the total 558) was selected. This approach allowed the models to train on a reduced but diverse feature set in each iteration, enhancing robustness while still leveraging the high-dimensional nature of radiomics data.

Class imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE) (Nitesh V. Chawla, 2002), which was applied within training folds when the minority class size was sufficient. SMOTE generates synthetic instances of the minority class by interpolating between existing samples, helping to balance the dataset and improve classification robustness. Cross-validation was performed using GroupKFold to

ensure that no images from the same patient were split across training and validation sets, thereby preventing data leakage. Model evaluation was based on multiple metrics, including accuracy, balanced accuracy, F1-score, and Matthews Correlation Coefficient (MCC). Finally, SHAP (SHapley Additive exPlanations) was applied to interpret model predictions and identify the most influential radiomic features across imaging modalities.

3.4 Evaluation Metrics

In order to compare the behaviour of the different models and assess their generalization performance, an evaluation was implemented (Maier-Hein, L., Reinke, A., Godau, P. et al., 2024). This includes multiple performance metrics (Accuracy, F1 Score, Balanced Accuracy, MCC, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC)) computed at three levels: cross-validation, confusion matrix-based predictions, and predictions on an independent dataset. This multi-level approach provides deeper insights into model reliability, performance under class imbalance, and feature consistency across training folds.

For the following definitions consider:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

3.4.1. Accuracy

Accuracy, as defined on Equation 1, measures the proportion of correctly classified instances among the total instances. While intuitive and easy to interpret, accuracy alone may be misleading in imbalanced datasets, where the model might perform well simply by predicting the majority class.

Equation 1: Formula for classification accuracy, representing the proportion of correct predictions over the total number of samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

3.4.2. F1 Score

The F1 Score, as defined on Equation 2, is the harmonic mean of precision and recall. It balances the trade-off between false positives and false negatives, making it more informative than accuracy when class distribution is uneven. This metric is especially important when both false negatives (e.g., missing a cancer diagnosis) and false positives (e.g., unnecessary interventions) carry significant consequences.

Equation 2: Formula for the F1 score, the harmonic mean of precision and recall, commonly used in imbalanced datasets.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall},$$

$$where\ Precision = \frac{TP}{TP + FP}\ , Recall = \frac{TP}{TP + FN}$$

3.4.3. Balanced Accuracy

Balanced Accuracy, as defined on Equation 3, accounts for imbalanced class distributions by averaging the recall (sensitivity) of each class. This ensures that both classes are given equal importance, providing a more realistic picture of model performance in cases where one class is underrepresented.

Equation 3: Balanced accuracy, computed as the average recall for both classes, correcting for class imbalance.

$$Balanced\ accuracy = \frac{1}{2} * \left(\frac{TP}{TP + EN} + \frac{TN}{TN + FP}\right)$$

3.4.4. Matthews Correlation Coefficient (MCC)

MCC, as defined on Equation 4, is a robust metric that considers true and false positives and negatives and is considered a balanced measure even in the presence of class imbalance. It produces a value between -1 and +1, where +1 indicates perfect prediction, 0 indicates no better than random, and -1 indicates total disagreement between prediction and observation.

Equation 4: MCC, a comprehensive performance metric that considers all values of the confusion matrix.

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

3.4.5. ROC Curve and AUC Score

The ROC curve plots the true positive rate (sensitivity) against the false positive rate at various threshold settings, showing how the model's performance varies with different classification thresholds. The Area Under the ROC Curve (AUC), as defined on Equation 5, condenses this performance into a single value, where a higher AUC indicates a better ability to distinguish between classes. This is particularly valuable in medical applications with imbalanced datasets, as it is threshold-independent and highlights discriminatory power.

Equation 5: AUC-ROC, representing the model's ability to distinguish between classes across all thresholds.

$$AUC - ROC = \int_0^1 TPR(FPR)dFPR\,,$$

$$where\ TPR = \frac{TP}{TP + FN}\ , FPR = \frac{FP}{FP + TN}$$

3.4.6. Feature Importance Across Folds

To gain interpretability, the top 10 most important features are identified for each cross-validation fold. This analysis reveals which features consistently influence model predictions, aiding both in model transparency and in understanding the biological or clinical relevance of certain predictors. Consistent feature importance across folds suggests stable and generalizable model behaviour.

3.4.7. Confusion matrix

A separate confusion matrix is presented for each trained model, enabling a direct comparison of how different approaches perform in identifying cancerous lesions. This facilitates a more informed selection of the most effective model for clinical application.
The confusion matrix illustrates the performance of the model in classifying breast cancer cases into two categories: benign and malignant. By comparing the model's predictions against the actual diagnoses, the confusion matrix provides a clear breakdown of true positives, true negatives, false positives, and false negatives.

Chapter 4

4. Implementation

4.1 Pipeline of the work

The pipeline illustrated in Figure 23 outlines a comprehensive approach to cancer classification (i.e., malignant vs benign) used in this work. Beginning with dataset acquisition, we implement a series of preprocessing steps including binary masking and image normalization. These steps are crucial for isolating regions of interest and ensuring consistent input features. The system extracts relevant features from the pre-processed images, combines them with clinical metadata, and feeds this information into a classifier model that has been carefully balanced, trained, and evaluated. As a result of all of these steps, the classifier outputs some plots where we can see its accuracy using different models, the confusion matrix and a comparison between the results of using DM or CESM. In order to do so, Table 3 and Table 4 describe the imports used in the code and the main classifier methods, respectively.

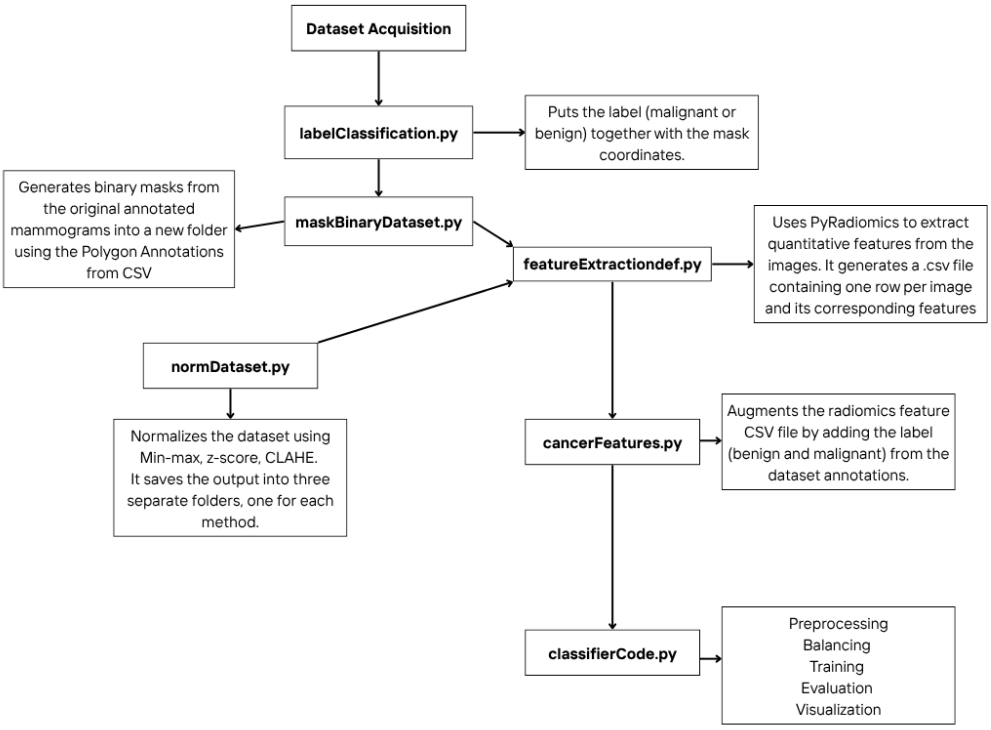

Figure 23: Pipeline of the python modules implemented in this work. Bold elements represent individual files, while arrows indicate what the file does.

Table 3: Python library imports used in the classifier script, organized by functional category.

Data handling and preprocessing	import numpy as np
	import pandas as pd
	from sklearn.preprocessing import LabelEncoder, StandardScaler
Model evaluation and metrics	from sklearn.metrics import accuracy_score, balanced_accuracy_score, f1_score, precision_score, recall_score
	from sklearn.metrics import matthews_corrcoef, cohen_kappa_score, log_loss
	from sklearn.metrics import roc_auc_score, roc_curve, auc, RocCurveDisplay
Cross-validation and model selection	from sklearn.model_selection import GroupKFold, RandomizedSearchCV, GridSearchCV
	from scipy.stats import randint, uniform
	from sklearn.utils import resample
Machine learning models	from sklearn.svm import SVC
	from sklearn.ensemble import RandomForestClassifier
	from xgboost import XGBClassifier
	from lightgbm import LGBMClassifier
	from sklearn.linear_model import LogisticRegression
	from sklearn.neighbors import KNeighborsClassifier
Dimensionality reduction	from sklearn.decomposition import PCA
Imbalanced data handling	from imblearn.over_sampling import SMOTE
Visualization and plotting	import matplotlib.pyplot as plt
	import seaborn as sns
	from IPython.display import Image, display
Explainability and interpretation	import shap

Table 4: Core methods implemented in the classifier script, with a description of their purpose and functionality. The functions described below belong to the classifierCode.py file, shown in the project pipeline overview in Figure 24, and are explained in the following the table.

Classifier's Methods	Task
_extract_features_and_labels(self)	Reads the CSV file that contains both the extracted radiomics features and their corresponding labels. It separates the input features from the labels and encodes the labels into a machine-readable format using a label encoder. It also extracts additional metadata, such as image filenames, which are later used to determine the imaging modality (DM or CESM), and patient IDs, which are crucial for splitting the dataset correctly during cross-validation.
preprocess_features(self, X_train, X_val=None, correlation_threshold=0.95, use_pca=False, pca_components=None)	Prepares the feature data for machine learning models. It standardizes all features to have zero mean and unit variance, which improves model convergence and stability. Optionally, it can apply PCA (Principal Component Analysis) to reduce dimensionality, keeping only the most significant components and discarding noise. It can also remove highly correlated features to simplify the feature space and prevent redundancy that could confuse learning algorithms.
_extract_patient_ids(self)	Extracts the unique patient identifier from the image filenames. This step is critical because medical images often include multiple views or scans from the same patient. By grouping images based on patient IDs, it ensures that cross-validation splits are based on patients, preventing data leakage between training and testing sets.
balance_dataset(self, X, y)	Addresses the common issue of class imbalance in medical datasets. It applies resampling techniques like SMOTE or SMOTETomek to balance the distribution of benign and malignant cases. This ensures that the model does not become biased toward the majority class and can learn to classify both classes effectively.
run_patient_based_cross_validation()	Manages the training and evaluation process across multiple folds. It splits the dataset based on patient IDs into training and validation sets, preprocesses features, applies balancing if needed, trains the selected models, and collects performance metrics like accuracy and confusion matrices. This method ensures that model evaluation is realistic and clinically meaningful by preventing patient overlap between training and testing sets.
compare_image_types()	Analyses the impact of different imaging modalities, DM and CESM, on model performance. It separates the dataset based on image type, trains models separately on each

	subset, and compares their classification accuracies. This provides insight into which imaging technique is more effective for detecting and classifying cancer types.
compare_models()	Trains and evaluates multiple machine learning models (including neural networks, support vector machines, random forests, and ensemble methods) on the same dataset. It collects and organizes the performance metrics, allowing for a direct comparison of how different types of models perform under the same conditions.
visualize_model_comparison()	Generates bar plots and box plots to compare the average performance of different models. These visualizations summarize and present the results in a clear, interpretable way, making it easy to see which model consistently performs best across cross-validation folds.
train_svm(self, X_train, y_train, X_val=None, y_val=None, tune_hyperparams=True)	Trains a SVM classifier. Optionally performs hyperparameter tuning using randomized search with cross-validation. It returns the best model and optionally evaluates its performance on a validation set.
train_random_forest(self, X_train, y_train, X_val=None, y_val=None, tune_hyperparams=True)	Trains a Random Forest classifier. Like the SVM method, it supports hyperparameter tuning and validation. Also computes and displays feature importances from the trained forest, aiding in interpretability.
train_xgboost(self, X_train, y_train, X_val=None, y_val=None, tune_hyperparams=True)	Trains an XGBoost classifier, optionally with randomized hyperparameter search. Evaluates the model on validation data and reports accuracy.
train_lightgbm(self, X_train, y_train, X_val=None, y_val=None, tune_hyperparams=True)	Trains a LightGBM classifier. Performs hyperparameter tuning if enabled. Like other training methods, it validates and reports model performance.
train_knn(self, X_train, y_train, X_val=None, y_val=None, tune_hyperparams=True)	Trains a kNN classifier. Performs grid search over different values of k to optimize performance. Evaluates the trained model on validation data.
train_logistic_regression(self, X_train, y_train, X_val=None, y_val=None, tune_hyperparams=True)	Trains a logistic regression model with optional hyperparameter tuning. Uses class balancing and different solvers depending on the hyperparameter configuration.
tune_hyperparameters(self, X_train, y_train, X_val, y_val)	Tunes neural network hyperparameters including learning rate, dropout rate, and batch size using grid search. Applies early stopping to avoid overfitting and uses SHAP to provide interpretability for the best models.
create_ensemble(self, X_train, y_train, X_val, y_val)	Creates an ensemble of different models (deep neural networks, SVM, Random Forest). Combines their predictions using soft-voting by averaging predicted probabilities. Evaluates individual models and the ensemble on validation data.
run_patient_based_cross_validation(...)	Performs cross-validation by grouping samples by patient, ensuring that data from the same patient does not appear

	in both training and validation sets. Trains and evaluates the selected model (or ensemble) across all folds and reports a comprehensive set of performance metrics, including accuracy, F1 score, MCC, and AUC. Supports modality-specific evaluations (DM vs CESM vs CESM+DM).
feature_importance(self, force_neural=False)	Estimates feature importance using either a Random Forest or a SHAP-based analysis with a neural network. Outputs a DataFrame ranking features by their impact on model predictions.
visualize_confusion_matrices(self, fold_results)	Plots the confusion matrix based on the predictions of the trained models. This visualization helps to quickly understand how well the model is performing, including how often it misclassifies benign tumours as malignant and vice versa.
plot_model_metric_by_image_type_2(self, model_results, metric)	Generates a grouped boxplot for a specified metric (F1, accuracy, MCC, balanced accuracy) across different models and image types (DM, CESM, ALL), providing insights into performance variance.

4.2 Generation of tumour’s binary mask

In the context of cancer classification from medical images, accurately isolating ROIs is a crucial step in ensuring that the features extracted are both relevant and informative. One powerful approach for this is the use of binary masks.

A binary mask is an image with the same dimensions as the original medical image, where each pixel is either 1 (included in the region of interest) or 0 (excluded). This mask acts as a filter that allows the analysis to focus only on specific areas identified as diagnostically significant, typically areas containing or surrounding tumours.

In this workflow, the ROIs are not arbitrarily defined. Instead, they are based on polygon coordinates provided by a medical expert, stored in a CSV file. Each set of coordinates outlines a specific area of interest, such as suspected malignant tissue, benign lesions, or other pathological regions.

Using binary masks in this way offers several advantages:

- Noise reduction: By ignoring the background and non-relevant tissues, we reduce the risk of extracting misleading or irrelevant features.
- Data efficiency: Feature extraction and model training become more efficient as the model focuses only on the most important areas.
- Improved accuracy: Machine learning models trained on features extracted from expert-annotated regions tend to perform better because the input data is more aligned with diagnostic criteria.
- Interpretability: It becomes easier to trace the model's predictions back to specific annotated regions, enhancing the interpretability and trustworthiness of the system.

To implement this, the CSV file containing the doctor’s annotations is used to extract the polygon coordinates. These polygons are then rendered onto a blank image to create the binary mask. This mask is applied to the original medical image to isolate the ROIs, and only the pixels within these regions are used for subsequent processing and feature extraction. (Figure 24).

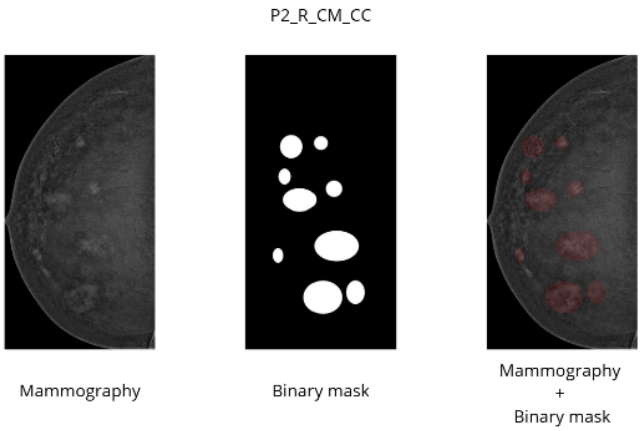

Figure 24: Example of binary mask on mammography.
The left image shows the CLAHE-normalized mammogram, the centre image displays the ROI mask, and the right image illustrates the overlay of the ROI mask on the normalized mammogram.

4.3 Image preprocessing

When extracting radiomic features using PyRadiomics, preprocessing steps play a crucial role in standardizing the input images and potentially enhancing relevant information. Min-Max scaling scales the intensity values to a specific range, typically 0 to 1, and is useful for ensuring all images have the same intensity range. However, it's sensitive to outliers, which can compress most of the data into a narrow range, and may not enhance subtle intensity variations within the image. Z-score normalization standardizes intensity values by subtracting the mean and dividing by the standard deviation, centring the data around zero with unit variance. Its limitations include the assumption of a Gaussian distribution of intensity values, which may not always be the case in medical images, and susceptibility to outliers, in addition to not enhancing local contrast. In contrast, CLAHE enhances local contrast by dividing the image into small tiles and applying histogram equalization to each tile, limiting the amplification of noise by clipping the histogram at a predefined value.

CLAHE offers several advantages for radiomics. Its ability to preserve and enhance local texture means that subtle intensity variations within specific regions of the image can lead to the extraction of more discriminative texture features. CLAHE also reduces intensity variability by enhancing contrast locally, which can make the extracted features more robust. Furthermore, it can reveal details that might be obscured by global normalization methods, potentially leading to the extraction of new or more distinct radiomic features.

Radiomics often relies on analysing the texture of an image, which refers to the spatial relationships between pixel intensities. Features that describe texture, like those from GLCM and GLRLM, quantify how often certain intensity patterns occur and how they are arranged. CLAHE enhances the contrast within small, local areas of the image, making subtle differences in intensity between neighbouring pixels more pronounced. Because

texture is based on these local intensity variations, CLAHE makes the texture within an image more visible and easier to quantify. By making the texture more apparent, CLAHE helps radiomic algorithms extract features that are more sensitive to the fine details of the image. If the texture is enhanced, the features extracted are more likely to be different from each other, which is important because distinct features provide more information for analysis and modelling. Min-Max scaling and Z-score normalization, on the other hand, primarily focus on adjusting the overall distribution of pixel intensities and don't specifically enhance the local variations that define texture. As a result, these methods were not as effective as CLAHE when highlighting the subtle textural differences that can be important in radiomics, and as a result, providing less features.

4.4 Features extraction

The starting point for this process was a dataset of mammographic images, each accompanied by a corresponding binary mask that delineated the ROI, typically a tumour or suspicious lesion. These masks were essential for focusing the feature extraction process on the relevant area. However, due to variability in naming conventions and dataset organization, a matching process was implemented to correctly associate each image with its mask. This included regex-based parsing of filenames to extract a shared identifier. The final image-mask pairs were loaded using the SimpleITK 2.5.0 library, which allowed for direct manipulation of medical imaging data.

Feature extraction was then performed using a custom YAML configuration file (Figure 25) that specified both preprocessing and feature calculation settings. Key settings included disabling image normalization (to preserve raw intensity distribution), setting a fixed bin width of 25 for intensity discretization, and forcing 2D analysis mode to align with the nature of the images. The configuration also defined which feature classes to compute and which image filters to apply. Features were extracted from the original images as well as from filtered versions, LoG and multi-resolution wavelet decompositions. These filtered images highlight structures at different spatial scales and frequencies, thereby enriching the feature space with multi-scale texture information.

```
def create_params_file(filename="params.yaml"):
    with open(filename, 'w') as f:
        f.write("""# PyRadiomics parameters file
setting:
  normalize: False
  normalizeScale: 100
  binWidth: 25
  # Fix for 2D images - use 2D resampling
  resampledPixelSpacing: [1, 1]
  interpolator: 'sitkBSpline'
  correctMask: True
  geometryTolerance: 1e-3
  force2D: True
  # Specify which dimension to use for 2D extraction
  # (0=axial, 1=coronal, 2=sagittal)
  force2Ddimension: 0

imageType:
  Original: {}
  Wavelet: {}
  LoG:
    sigma: [1.0, 3.0, 5.0]

featureClass:
  firstorder: []
  glcm: []
  glrlm: []
  glszm: []
  gldm: []
  ngtdm: []
""")
```

Figure 25: Screenshot of the YAML configuration file used by PyRadiomics for image preprocessing for radiomic features extraction.

After extracting features from all available image-mask pairs, the results were compiled into a structured dataset, with each row corresponding to an individual image and each column representing a specific radiomic feature. The final dataset contained around five hundred features per image.

This dataset was then saved to a CSV file for downstream analysis. Importantly, each image retained its unique identifier, allowing for patient-level grouping and stratified evaluation across different image types. In doing so, this radiomics pipeline transformed raw medical images into structured, high-dimensional numerical data suitable for machine learning, while preserving anatomical specificity and clinical interpretability.

4.5 Machine learning for radiomics-based classification

Once radiomics features were extracted from both DM and CESM, they were used as the base for training the machine learning algorithms designed to classify breast lesions as benign or malignant. A classification pipeline was designed not only to optimize classification performance but also to ensure clinical validity through cross-validation, preprocessing and interpretability. The process includes data transformation, model training, evaluation, and ensemble learning, with a special emphasis on avoiding data leakage and accounting for patient-specific variability.

Before any model training could occur, the radiomics dataset was subjected to several preprocessing steps. First, all features were standardized using z-score normalization. This step was critical, particularly for algorithms that are sensitive to feature scale, such as SVM and NN. Standardization ensured that all features contributed equally during training and prevented any single high-magnitude feature from dominating the learning process. Subsequently, highly correlated features were removed using a correlation threshold of 0.95. This form of redundancy reduction minimized the risk of multicollinearity, which can impair model interpretability and inflate variance in decision boundaries. In selected experiments, Principal Component Analysis (PCA) was applied as an optional dimensionality reduction technique, enabling compression of the feature space while preserving most of its variance.

Handling class imbalance was another key component of the pipeline. The dataset exhibited an uneven distribution of benign and malignant cases, which is a common issue in medical imaging datasets. To address this, SMOTE was integrated. SMOTE generates synthetic samples of the minority class by interpolating between existing samples. However, SMOTE was used selectively; it was activated only when the minority class contained enough samples to produce reliable synthetic data. This safeguard prevented the introduction of noise when the sample size was critically low, ensuring that model performance was not artificially inflated by poorly representative synthetic instances.

One of the core principles of this framework was the use of patient-based cross-validation. To ensure robust evaluation and prevent data leakage, the dataset was split using patient-based cross-validation via GroupKFold, with patient identifiers extracted from filenames serving as the grouping variable. The data was split into a training set and a validation (test) set based on patient grouping. All radiomic features and corresponding labels from a given patient were included in only one of these subsets per fold. The training set was used for model fitting, feature selection, class balancing, and, when enabled, hyperparameter tuning. The validation set was used exclusively for model evaluation and remained untouched during training to provide an unbiased estimate of performance. This guaranteed that all images from the same patient were assigned to either the training or the validation set within a fold, never both, thereby preserving the independence of samples across folds. This ensured that all images from a single patient were either in the training set or the validation set but never separated across both.

Each model was trained and evaluated within this cross-validation scheme, with hyperparameter tuning performed using either GridSearchCV or RandomizedSearchCV. Parameter grids were predefined for each algorithm, and tuning was carried out within the training data of each fold to avoid biasing performance estimates. During training, a feature resampling strategy was also employed to manage the high dimensionality of the radiomic feature space.

Learning strategies were tailored to the type of model. For classical machine learning models (e.g., SVM, Random Forest, Logistic Regression, kNN, XGBoost, and LightGBM), standard training procedures were applied with attention to regularization and feature scaling. For deep learning models built using Keras, batch normalization and dropout layers were used to stabilize and regularize training.

A wide array of machine learning algorithms was explored. Classical models included SVM, RF, Logistic Regression, kNN, XGBoost, and LightGBM. Each algorithm was evaluated with and without hyperparameter tuning. For tuning, randomized or grid search was used with cross-validation over predefined parameter spaces. In parallel, feedforward neural networks using TensorFlow/Keras were developed. These deep learning models incorporated batch normalization and dropout layers to mitigate overfitting and improve generalization. Architectures varied in depth and width, with tuning focused on learning rate, dropout rate, and batch size.

To leverage the strengths of individual models while mitigating their weaknesses, an ensemble classifier was used. This ensemble included two distinct neural networks, an SVM, and a Random Forest. Predictions from all models were aggregated using soft voting, where class probabilities were averaged to determine the final prediction. This approach helped to reduce model variance and improve robustness, especially on difficult or borderline cases.

Model evaluation was comprehensive and multifaceted. For each fold in cross-validation, traditional metrics were computed, such as accuracy and F1-score, but also included balanced accuracy and MCC to account for class imbalance. Confusion matrices were aggregated across folds to visualize classification patterns, and AUC-ROC was used to assess discriminative power. Precision-recall curves, Cohen's Kappa, and log loss were also calculated to provide a deeper understanding of classifier behaviour.

To ensure transparency and interpretability, SHAP was added to the trained models. SHAP summary plots and bar plots allowed to identify which radiomic features had the most significant impact on classification decisions. These insights are especially valuable in medical contexts, where understanding why a model makes a certain prediction is as important as the prediction itself.

Finally, performance was analysed not only across the entire dataset but also within subsets based on image type (DM alone vs CESM alone vs CESM+DM). This stratified evaluation revealed differences in model performance across imaging modalities and helped to identify which image type performed better across the multiple models. The comparison was further reinforced through visualizations of ROC curves, bar plots of different metrics, and statistical testing using paired t-tests between models.

Chapter 5

5. Results and discussion

5.1 Model and image type comparison

To determine the most effective approach for classifying breast cancer cases, the code evaluates and compares the performance of eight different machine learning models: Neural Network, SVM, Random Forest, Ensemble Method, XGBoost, Logistic Regression, kNN, and LightGBM.

The comparison is based on key performance metrics, such as accuracy, balanced accuracy, MMC, and F1-score. By analysing these metrics side by side, we aim to identify which model offers the best balance between correctly detecting malignant cases and minimizing false positives.

The plots below are boxplots, which visually summarize the distribution of performance metrics across all cross-validation folds for each model. These plots help highlight variability, consistency, and potential outliers in model performance. Specifically:

- Box: Represents the interquartile range (IQR), covering the middle 50% of the results, values between the 25th and 75th percentiles.
- Middle line inside the box: Indicates the median value across folds, providing a robust central estimate of model performance.
- Whiskers: Extend to the minimum and maximum values excluding outliers, showing the typical range of performance.
- Dots outside the whiskers: Represent outliers, or individual folds where the model performed significantly better or worse than usual.

These boxplots allow for quick visual comparison between models, revealing not only average performance but also stability and sensitivity to the data split.

Figure 26 shows the distribution of classification accuracies across the different models, stratified by image type (DM, CESM, and ALL). It is evident that models trained exclusively using CESM features consistently outperformed those trained with DM or the combined features. Among the classifiers evaluated, the Ensemble, LightGBM, and XGBoost models demonstrated the highest median accuracies and lowest variability when trained with CM. This result reinforces the hypothesis that radiomic features derived from CESM images are more informative for lesion classification than those from DM or mixed modalities.

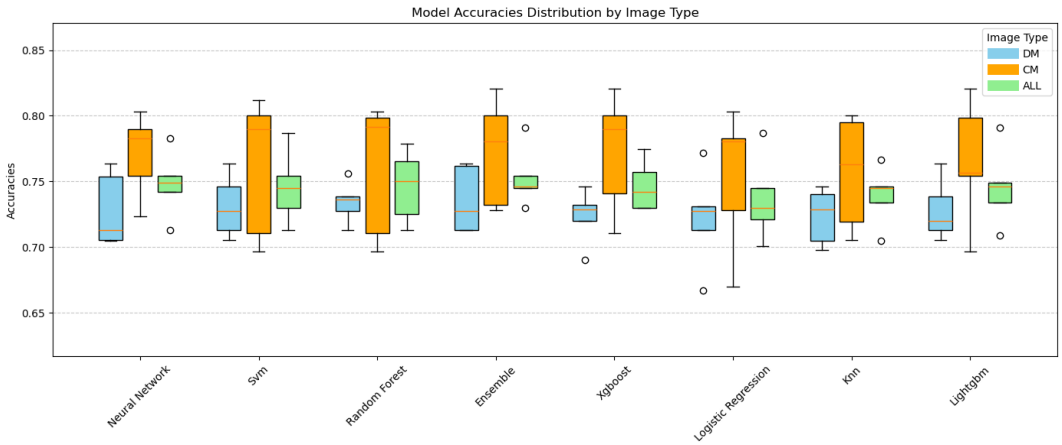

Figure 26: Model accuracies distribution by image type across all the models as a result of the X-validation test.

In Figure 27 a similar trend when evaluating balanced accuracy. As in the previous analysis based on overall accuracy, models trained with CESM outperformed those trained with DM or the combined features across nearly all classifiers. Notably, XGBoost, LightGBM and Ensemble achieved the highest balanced accuracy scores when trained exclusively with CESM images.

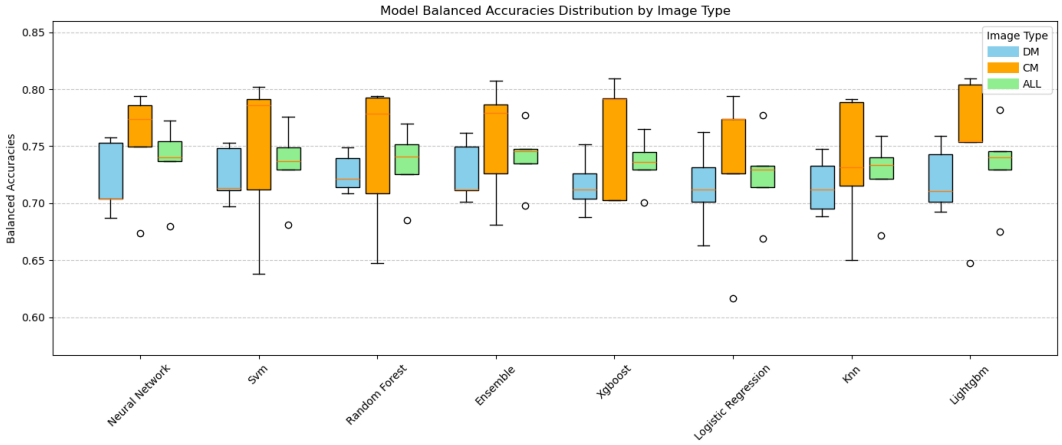

Figure 27: Model balanced accuracies distribution by image type across all the models as a result of the X-validation test.

Figure 28 displays the distribution of MCC scores across all models and image types. MCC is a balanced metric that considers all elements of the confusion matrix, making it especially suitable for imbalanced classification problems. As observed with accuracy and balanced accuracy, models trained with CESM features significantly outperformed those trained on DM or the combined features. In particular, LightGBM, Ensemble, and XGBoost achieved the highest median MCC scores with CESM data.

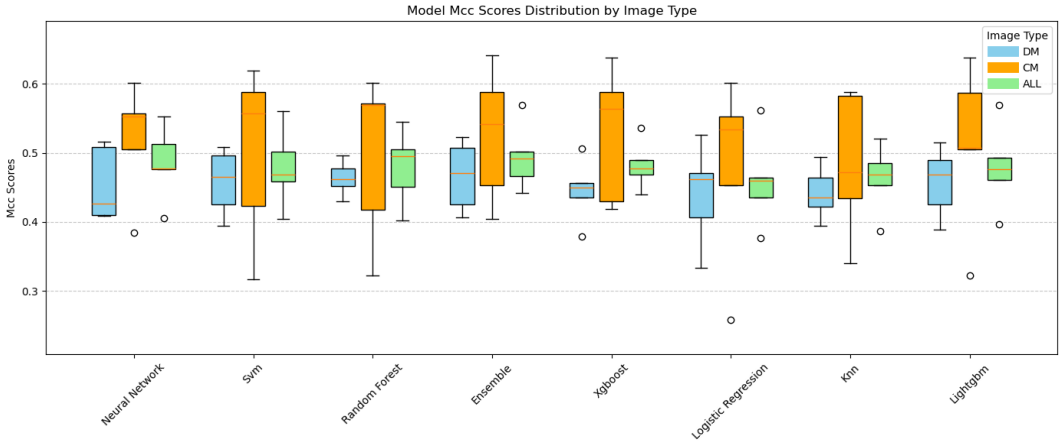

Figure 28: Model MCC scores by image type across all the models as a result of the X-validation test.

In Figure 29 a similar trend when evaluating F1 Score. As in the previous analysis based, models trained with CESM outperformed those trained with DM or the combined features across nearly all classifiers. XGBoost and Ensemble achieved the highest balanced accuracy scores when trained exclusively with CESM images.

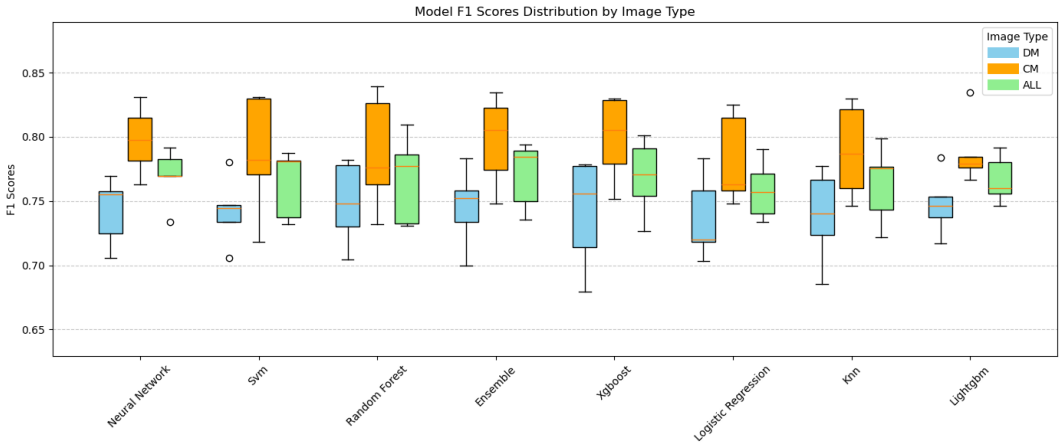

Figure 29: Model F1 Scores by image type across all the models as a result of the X-validation test.

5.2 Individual models performance

For each model, the code provides a comprehensive evaluation across all folds of the cross-validation process. Rather than only reporting overall performance model comparison at the end of training and testing, it outputs detailed metrics for each fold and at the end of each model, offering better understanding of model consistency and robustness.

Also, individually, it displays metrics such as Accuracy, F1 Score, Balanced Accuracy, and MCC at three levels: the full cross-validation summary, predictions made using the

confusion matrix, and predictions on a distinct dataset. This level of detail enables deeper insights into the generalization ability of each model.

Additionally, the AUC-ROC score is computed and the ROC curve is plotted for each model, in order to see the difference on the different types of images. The ROC curve illustrates the trade-off between the true positive rate (sensitivity) and the false positive rate across different threshold values, while the AUC score summarizes this performance into a single value. A higher AUC indicates better model ability to distinguish between classes, making it a valuable measure, especially in imbalanced classification tasks. (Figure 30).

Furthermore, the code also highlights the top 10 most important features for each fold, allowing us to examine how consistently certain features contribute to classification across different training subsets. This helps in identifying which features are most influential in the decision-making process of the models. (Table 5).

As shown in Figures 26-29, the XGBoost model demonstrated one of the best performances metrics (Table 6) across all folds compared to the other models. While all models were evaluated using the same metrics and graphical representations, XGBoost is presented here as a representative example due to its superior results. The following plots and tables illustrate key evaluation metrics observed during the training and validation process, highlighting the model's robustness and consistency.

Table 5: Top 10 features by importance output for fold 5 using XGBoost model.

Feature importance	Name of feature
0.022716	original_glrlm_LongRunEmphasis
0.018586	log-sigma-1-0-mm-3D_ngtdm_Strength
0.018168	original_gldm_DependenceNonUniformity
0.017149	wavelet-H_ngtdm_Coarseness
0.015457	log-sigma-1-0-mm-3D_glcm_ClusterShade
0.014635	log-sigma-5-0-mm-3D_ngtdm_Coarseness
0.014565	original_firstorder_Energy
0.013785	log-sigma-1-0-mm-3D_firstorder_Median
0.012680	log-sigma-3-0-mm-3D_ngtdm_Contrast
0.012301	original_gldm_DependenceNonUniformityNormalized

The metrics showed on Table 6 correspond to the overall model's performance:

Table 6: Cross validation summary on DM, CESM and ALL image types, the higher values are highlighted.

	DM	CESM	ALL
Accuracy	0.7189	**0.7706**	0.7432
Balanced accuracy	0.7108	**0.7528**	0.7299
MCC	0.4343	**0.5189**	0.4736
F1 Score	0.7388	**0.8001**	0.7684
AUC Score	0.7180	**0.7632**	0.7384

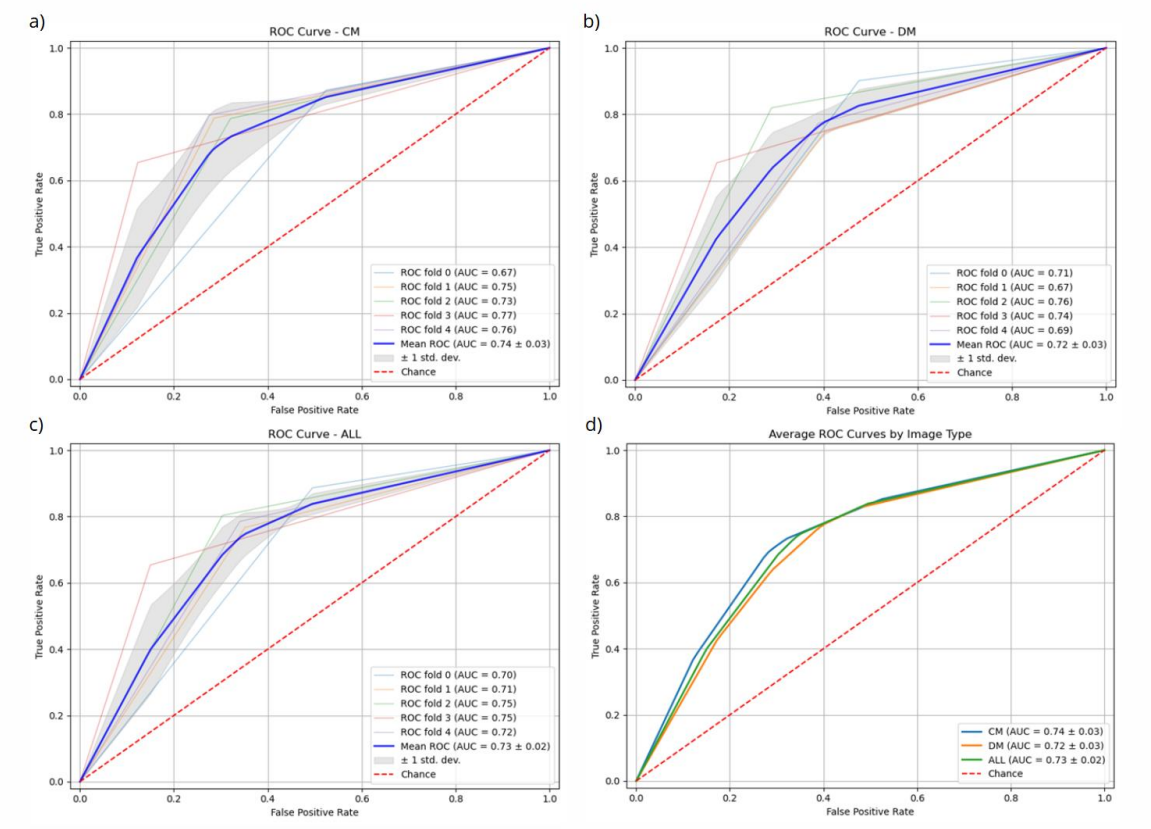

Figure 30: ROC curves for the XGBoost model.
(a) ROC curve for CESM-derived features. (b) ROC curve for DM-derived features.
(c) ROC curve for the combined feature set (CESM + DM). (d) Overlay of the mean ROC curves from all three modalities for direct comparison.

The ROC curves displayed in Figure 30 illustrate the performance of the XGBoost model across multiple cross-validation folds. Each curve represents the trade-off between

sensitivity (true positive rate) and specificity (false positive rate) for a different fold, showing how the model distinguishes between the classes under varying classification thresholds. The closer the curve follows the top-left corner, the better the model's overall discriminative ability.

In Figure 31, the diagonal elements (top-left and bottom-right) represent correct classifications: 393 benign cases were correctly identified as benign, and 514 malignant cases were correctly classified as malignant. The off-diagonal elements represent misclassifications: 177 benign cases were incorrectly labelled as malignant (false positives), while 135 malignant cases were mistakenly classified as benign (false negatives). These results demonstrate the model's ability to distinguish between cancer types, while also highlighting areas for potential improvement, particularly in reducing false positives.

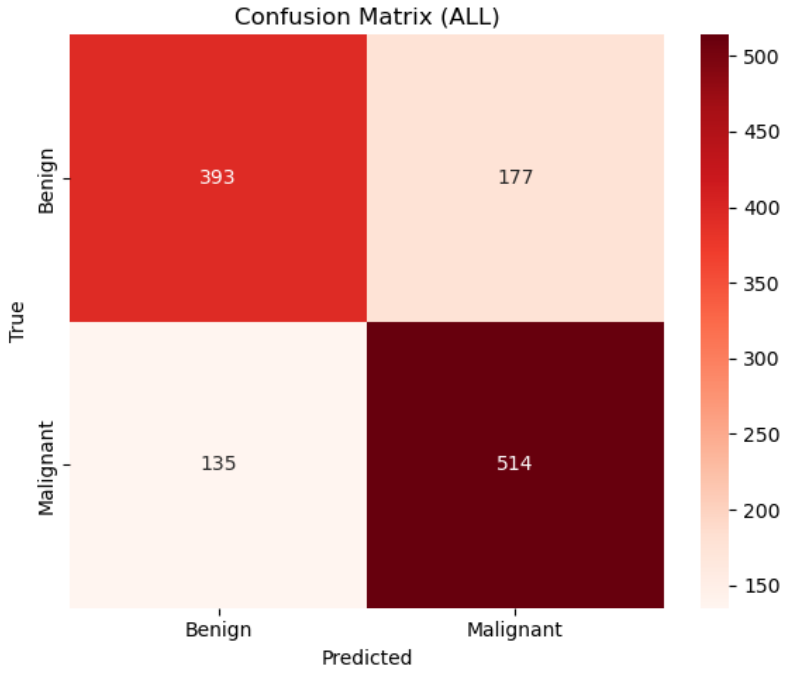

Figure 31: Confusion matrix using XGBoost model for all images.

In Figure 32, the diagonal elements (top-left and bottom-right) represent correct classifications: 179 benign cases were correctly identified as benign, and 255 malignant cases were correctly classified as malignant. The off-diagonal elements represent misclassifications: 70 benign cases were incorrectly labelled as malignant (false positives), while 68 malignant cases were mistakenly classified as benign (false negatives).

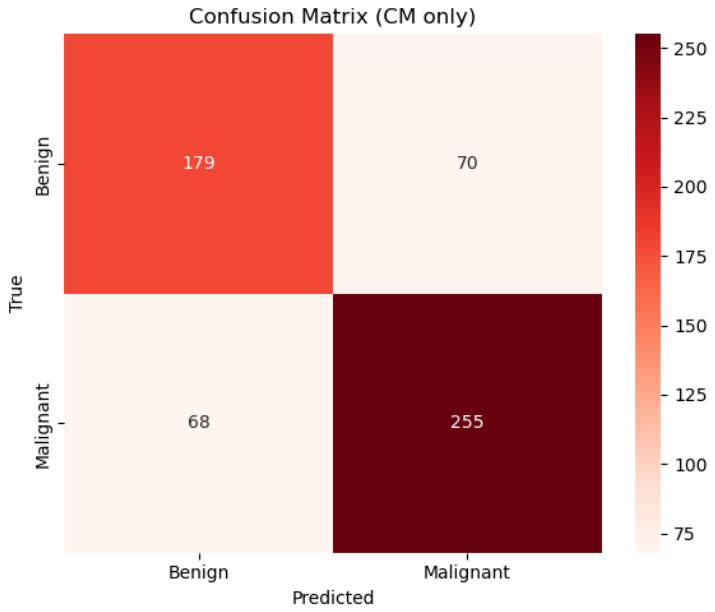

Figure 32: Confusion matrix using XGBoost model for CESM images.

In Figure 33, the diagonal elements (top-left and bottom-right) represent correct classifications: 214 benign cases were correctly identified as benign, and 259 malignant cases were correctly classified as malignant. The off-diagonal elements represent misclassifications: 107 benign cases were incorrectly labelled as malignant (false positives), while 67 malignant cases were mistakenly classified as benign (false negatives).

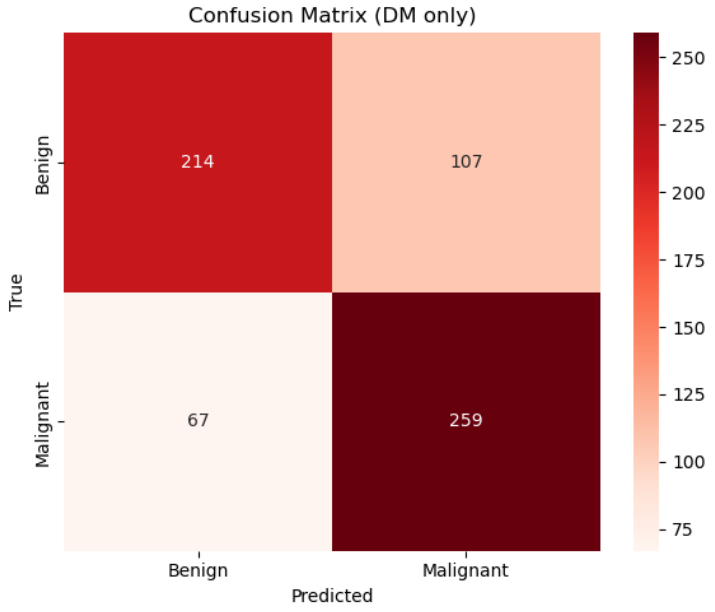

Figure 33: Confusion matrix using XGBoost model for DM images.

5.3 Feature importance by image type

In order to gain deeper insights into the decision-making process of the trained models, there was performed an explainability analysis using SHAP, a model-agnostic method for interpreting complex predictions. SHAP assigns a value to each feature based on its contribution to the final output, which allows to identify which radiomic features are strongly influencing the classification outcomes. To make a comparison on the different image types, SHAP analysis was conducted separately for each image type (CESM, DM and both combined). This separation enabled to explore whether different features are prioritized by the model depending on the imaging modality, and whether certain radiomic patterns are more predictive in one context than the other. The following results present SHAP value summaries and feature rankings for model LightGBM as an example, highlighting key differences in feature importance across image types (Figures 34-36).

Figure 34 is SHAP summary plot illustrates the top 25 radiomic features contributing to the model's classification decisions when trained exclusively with CESM features. The most impactful feature is original_glrlm_RunEntropy, indicating that run-length matrix-based heterogeneity in pixel intensities plays a key role in malignancy prediction. The SHAP value range spans approximately -0.5 to +0.3, reflecting moderate variability in how individual feature values affect predictions.

The dominance of GLRLM features is apparent, with LongRunEmphasis, GrayLevelNonUniformity, and LowGrayLevelRunEmphasis all contributing significantly. Several filtered features (e.g., LoG and Wavelet) also appear, suggesting that texture transformations add discriminative power.

The spread and clustering of SHAP values across features indicate that both high and low values of these features can push predictions in different directions, confirming the model's nuanced use of radiomics for CESM data.

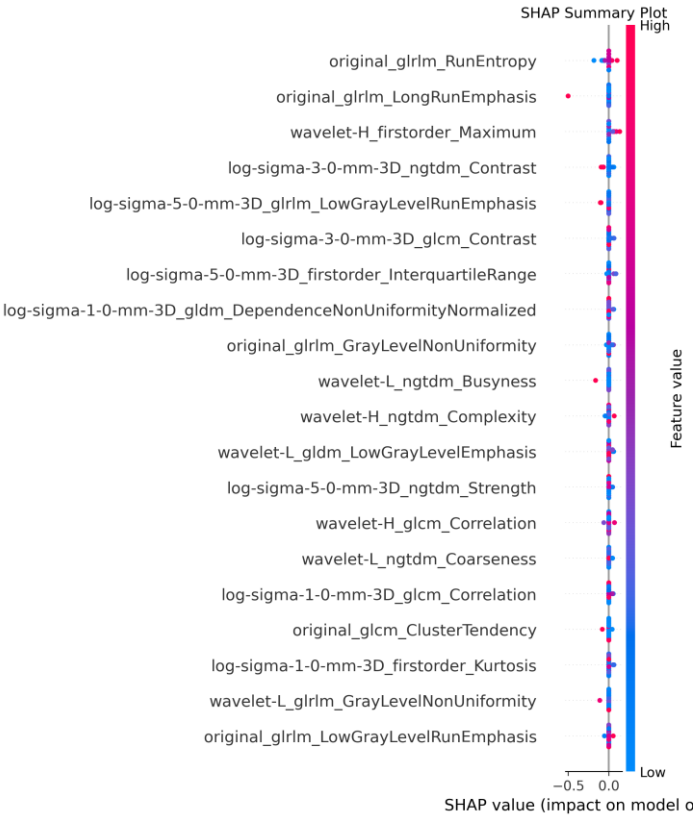

Figure 34: SHAP summary plot showing the most influential radiomic features for the LightGBM trained with CESM features. The feature original_glrlm_RunEntropy had the greatest overall impact on model output, highlighting the importance of texture heterogeneity in CESM-based classification.

As shown on Figure 35, for DM features, the most relevant feature is log-sigma-3-0-mm-3D_glcm_MaximumProbability, a GLCM feature derived from Laplacian of Gaussian-filtered images. This suggests that the model heavily relies on texture uniformity patterns for classification in the absence of contrast enhancement. Compared to CESM images, the SHAP values here span a more constrained range (about -0.25 to +0.15), implying more stable and less extreme feature contributions.

The feature set is largely dominated by GLCM and first-order features, with very few GLRLM or GLDM features ranking highly. This narrower feature importance profile may reflect the lower discriminative richness of DM images, as fewer modalities and intensity dynamics are available. The tighter clustering also suggests a more cautious model with less variability in how features influence predictions.

Figure 35: SHAP summary plot for LightGBM trained with DM features. The most impactful feature was log-sigma-3-0-mm-3D_glcm_MaximumProbability, indicating that local intensity relationships dominate the model's decision-making in the absence of contrast enhancement.

When combining both CESM and DM image features, the model again ranks original_glrlm_RunEntropy as the most influential, reinforcing its robustness across modalities. Interestingly, the SHAP value range here is the most constrained (approximately -0.1 to +0.1), suggesting that while the model is highly sensitive to small changes, the magnitude of impact per feature is smaller, likely due to feature dilution or redundancy from combining two modalities. (Figure 36).

A more diverse feature set appears in this plot, with contributions from GLRLM, first-order statistics (e.g., Skewness, Energy), and GLDM features. This variety reflects the attempt to integrate complementary information, but also leads to wider spreads for top features, indicating potential interactions or overlaps between CESM and DM contributions.

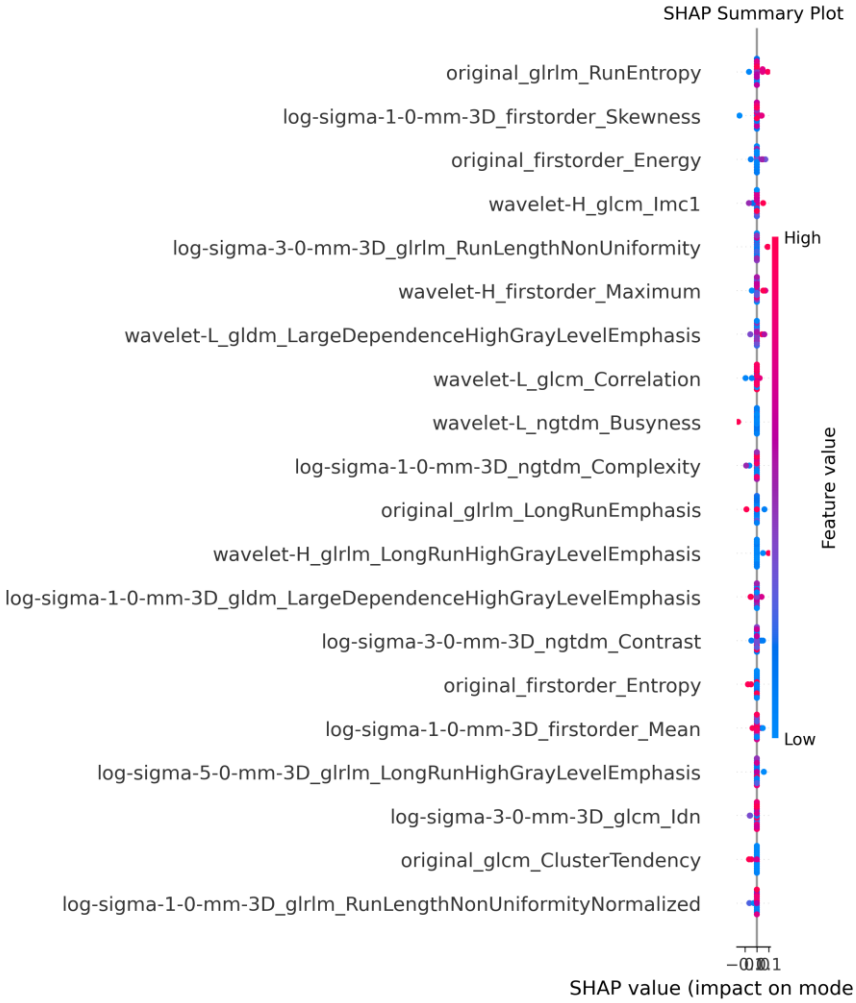

Figure 36: SHAP summary plot for the LightGBM model trained on the combined CESM and DM radiomic features. While original_glrlm_RunEntropy remained the most influential feature (highlighting the continued importance of CESM-derived texture) other features contributed more evenly across categories. However, the SHAP value distribution was narrower and more diluted than in the CESM-only model, suggesting that combining modalities introduces feature redundancy and reduces the individual impact of key predictors.

These results emphasize the impact of contrast enhancement on the nature and interpretability of radiomic features. In CESM images, the dominance of GLRLM features suggests that patterns related to run-length variability and tissue heterogeneity become more prominent and diagnostically useful. In DM, where such contrast is absent, the model instead relies more on local intensity relationships captured by GLCM features. When both modalities are combined, the model integrates a wider set of features, indicating a more complex interaction of texture descriptors across image types. This suggests that CESM not only improves overall classification performance, as previously shown, but also alters the radiomic signature of malignant versus benign lesions in a way that influences which features are most relevant.

Chapter 6

6. Conclusions and future work

6.1 Conclusions

The central hypothesis of the thesis was that CESM would provide more powerful radiomic features, and therefore higher diagnostic value, than those extracted from DM. This assumption was based on the different advantages of CESM in highlighting vascularization, contrast uptake, and lesion borders, factors that are especially important for differentiating between benign and malignant tissue *(Claudia Lucia Piccolo, 2024)*. It was expected that these characteristics would be reflected in the extracted radiomic features, making them more informative for training machine learning classifiers. Consequently, the hypothesis was that models trained on CESM images would perform better than those trained on DM images in terms of classification accuracy, robustness, and overall diagnostic reliability.

To evaluate this hypothesis, the images were evaluated separately using the same pipeline, using the most common machine learning algorithms and patient-level cross-validation. The results consistently confirmed the hypothesis: across all models, whether traditional models such as SVM and Random Forest or deep learning-based models such as neural networks, those trained with CESM images provided an overall 5-8% superior performance over DM (similar results to the works explained on the literature review), specially XGBoost, that showed one of the best results, including an increase in 7.19% accuracy, 5.91% balanced accuracy, 19.48% MCC and 8.3% F1-score. Even when models were trained on the combined dataset containing both image types, the performance remained lower than when trained solely on CESM, suggesting that including DM may introduce more noise or less discriminative information into the feature space.

These findings strongly validate the hypothesis and support that CESM imaging leads to the extraction of richer and more diagnostically relevant radiomic features. The improvement in classification performance is evident not only in raw metrics but also in visualizations such as SHAP plots, where CESM models showed clearer and more consistent patterns of feature importance. From a clinical perspective, this result proves the potential of CESM as a superior modality for radiomics-based diagnostic tools. It also highlights the importance of considering the different possible imaging modality when designing machine learning pipelines for medical image analysis.

6.2 Limitations

While the results of this thesis support the use of radiomics and machine learning for breast cancer classification, particularly with contrast-enhanced mammography, it is important to acknowledge several limitations that affected the scope and depth of it.

Firstly, the study was constrained by time and computational resources, which limited the scale of model tuning, the exploration of more computationally intensive deep learning architectures, and the ability to perform extensive external validations. For example, by doing SHAP values analysis, not all the features were considered because of memory shortage.

Secondly, the analysis was based on a single dataset (CDD-CESM Dataset), which limits the generalizability of the findings. The lack of multi-centre data introduces potential biases due to acquisition protocols, imaging hardware and population characteristics. Without external validation on independent datasets, the robustness of the models in broader clinical settings remains uncertain. Despite balancing classes through techniques such as SMOTE, the relatively small sample size restricts the complexity of the models and increases the risk of overfitting.

Finally, although SHAP analysis was used to understand which radiomic features had the most influence on the model's predictions, but due to time, no comparison was made between these features and established clinical knowledge. In other words, while the models highlighted certain features as important for distinguishing benign from malignant lesions, I could not verify whether those features correspond to the characteristics that radiologists typically use when evaluating breast images, such as lesion shape, margins, or density. As a result, even though the models are interpretable from a technical standpoint, their outputs have not yet been validated against expert clinical understanding. Further work would be needed to determine whether the most predictive features identified by the model are also meaningful and recognizable to clinicians in a real diagnostic setting.

6.3 Future work

While this thesis highlights the important role of radiomics-based machine learning in classifying breast lesions, there are still some important areas for further investigation that could enhance the clinical utility of such approaches. A critical next step involves systematically comparing the performance of these radiomics-driven models with conventional diagnostic practices. For instance, juxtaposing the predictive accuracy of machine learning algorithms with assessments by experienced radiologists or standardized clinical tools such as the BI-RADS would help quantify the incremental benefit that radiomics may bring to diagnostic workflows. This comparison is essential for

understanding whether these advanced computational methods can meaningfully support or even outperform human judgment in routine clinical settings.

Beyond model comparison, another promising avenue lies in the integration of radiomic features with a broader spectrum of clinical data. Patient demographics (such as age or menopausal status), tumour biology (like hormone receptor expression or HER2 status), and histopathological findings could be incorporated alongside image-derived features to build more holistic predictive models. This multimodal approach could significantly improve the ability to subtype breast cancer more precisely, assess individual risk profiles, and anticipate treatment responses. Such enriched models would have greater potential for aiding in personalized treatment planning and improving patient outcomes.

To ensure that the findings are generalizable and not confined to a specific dataset or institutional setting, future studies should prioritize external validation. Applying the trained models to datasets from other sources would provide a more rigorous assessment of their performance under real-world variability.

Another exciting direction involves the combination of handcrafted radiomics with deep learning-based feature extraction. While traditional radiomics relies on predefined image characteristics such as texture, shape, and intensity, deep learning, especially through Convolutional Neural Networks (CNNs), can automatically learn hierarchical and abstract representations from raw image data. Integrating these complementary approaches could capture a broader range of image features, potentially leading to more accurate and nuanced diagnostic tools.

Finally, expanding beyond classification tasks to include prognostic modelling represents a natural evolution of this work. By incorporating time-to-event data, such as time to recurrence or survival duration, machine learning models could shift from merely identifying malignancies to predicting long-term patient outcomes. This capability would be crucial for designing risk-adaptive follow-up strategies and guiding treatment decisions over the course of a patient's disease trajectory, ultimately contributing to more personalized and effective healthcare.

Bibliography and References

American Association for Cancer Research. (2010). Retrieved from Genetic and epigenetic alterations in cancer: Implications for diagnosis and therapy.: https://pmc.ncbi.nlm.nih.gov/articles/PMC2804038/

American Cancer Society. (n.d.). Retrieved from Breast cancer survival rates: https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html

American College of Radiology. (2013). Retrieved from BI-RADS: https://www.acr.org/Clinical-Resources/Clinical-Tools-and-Reference/Reporting-and-Data-Systems/BI-RADS

American Family Physician. (2018). Retrieved from Breast Cancer Screening with MRI: More False-Positives, More Biopsies: https://www.aafp.org/pubs/afp/issues/2018/1115/od5.html

American Society of Clinical Oncology (ASCO). (2023). Retrieved from Journal of Clinical Oncology, 41(16_suppl), Article 10528. : https://ascopubs.org/doi/10.1200/JCO.2023.41.16_suppl.10528

Borghild Løyland, I. H. (2024). Retrieved from Causes and Risk Factors of Breast Cancer, What Do We Know for Sure? An Evidence Synthesis of Systematic Reviews and Meta-Analyses: https://www.mdpi.com/2072-6694/16/8/1583

Cancer Imaging Archive. (n.d.). Retrieved from Collection: CDD-CESM.: https://www.cancerimagingarchive.net/collection/cdd-cesm/

Cancer Research UK. (2021). Retrieved from Survival for breast cancer: https://www.cancerresearchuk.org/about-cancer/breast-cancer/survival

Claudia Lucia Piccolo, M. S. (2024). Retrieved from Radiomics for Predicting Prognostic Factors in Breast Cancer: Insights from Contrast-Enhanced Mammography (CEM): https://www.mdpi.com/2077-0383/13/21/6486

Cleveland Clinic. (2024). Retrieved from Breast Biopsy: https://my.clevelandclinic.org/health/diagnostics/24204-breast-biopsy-overview

Corey Whelan . (2025). Retrieved from Breast Biopsy Procedure: What to Expect: https://www.verywellhealth.com/breast-biopsy-7966144

Debbi, K. M.-L. (2023). Retrieved from Radiomics model to classify mammary masses using breast DCE-MRI compared to the BI-RADS classification performance.: https://doi.org/10.1186/s13244-023-01404-x

Diaz, O. K. (2021). Retrieved from A comprehensive guide to open-access platforms and tools. Physica medica, 83, 25-37.: https://www.sciencedirect.com/science/article/pii/S1120179721000958

Dr. Chan Ching Wan. (2021). Retrieved from What if Something is Detected on My Mammogram?: https://www.mountelizabeth.com.sg/health-plus/article/what-if-something-is-detected-on-my-mammogram

Dr. Mary Ling. (n.d.). Retrieved from Breast anatomy: https://www.drmaryling.com.au/breast-anatomy

Ergul, Nurhan & Kadioglu, Huseyin & Yildiz, Seyma & Yucel, Serap & Gucin, Zuhal & Erdogan, Ezgi & Aydin, Mehmet & Muslumanoglu, Mahmut. (2014). Retrieved from Assessment of multifocality and axillary nodal involvement in early-stage breast cancer patients using 18F-FDG PET/CT compared to contrast-enhanced and diffusion-weighted magnetic resonance imaging and sentinel node biopsy.: https://www.researchgate.net/publication/263862228_Assessment_of_multifocality_and_axillary_nodal_involvement_in_early-stage_breast_cancer_patients_using_18F-FDG_PETCT_compared_to_contrast-enhanced_and_diffusion-weighted_magnetic_resonance_imaging_and

European Commission. (2021). Retrieved from EU Missions in Horizon Europe: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe_en

Francesca Gallivanone,Gloria Bertoli and Danilo Porro. (2022). Retrieved from Radiogenomics, Breast Cancer Diagnosis and Characterization: Current Status and Future Directions.: https://www.mdpi.com/2409-9279/5/5/78

GitHub user @katjagerlach. . (2023). Retrieved from Error while trying to use a parametric map in Pyradiomics [Issue #828]. GitHub. : https://github.com/AIM-Harvard/pyradiomics/issues/828

Hussain, S., Qureshi, H. N., Qazi, H. J., & Saeed, M. . (2023). Retrieved from Deep learning, radiomics and radiogenomics applications in the digital breast tomosynthesis: A systematic review. : https://doi.org/10.1186/s12859-023-05515-6

Image Biomarker Standardisation Initiative. . (2023). Retrieved from Image Biomarker Standardisation Initiative. : https://theibsi.github.io/

Janusz Skowronek. (2011). Retrieved from CT-Image Guided Brachytherapy: https://www.researchgate.net/publication/221911642_CT-Image_Guided_Brachytherapy

Jou, S. R. (2019). Retrieved from An entropy-based framework for detecting cancer cell morphology changes. Entropy, 21(11), 1110. : https://www.mdpi.com/1099-4300/21/11/1110

Khaled, R., Helal, M., Alfarghaly, O. et al. (2021). Retrieved from Categorized contrast enhanced mammography dataset for diagnostic and artificial intelligence research: https://www.nature.com/articles/s41597-022-01238-0?utm_source=chatgpt.com

Lambin, P. L. (2017). Retrieved from Radiomics: The bridge between medical imaging and personalized medicine. : https://doi.org/10.1038/nrclinonc.2017.141

Lemoine, C. L. (2023). Retrieved from Introduction to radiomics and AI in radiology: A beginner's guide. Insights into Imaging, 14, Article 70. : https://doi.org/10.1186/s13244-023-01415-8

Liang, Y., Xu, H., Lin, J. et al. . (2025). Retrieved from Multi-modal radiomics model based on four imaging modalities for predicting pathological complete response to neoadjuvant treatment in breast cancer: https://bmccancer.biomedcentral.com/articles/10.1186/s12885-025-14407-2

Losurdo, L. F. (2019). Retrieved from Radiomics Analysis on Contrast-Enhanced Spectral Mammography Images for Breast Cancer Diagnosis: A Pilot Study: https://www.mdpi.com/1099-4300/21/11/1110

Lundberg, S. M.-I. (2023). Retrieved from Explaining machine learning models and predictions: An introduction. Nature Methods, 20, 277–288: https://doi.org/10.1038/s41592-023-02151-z

Maier-Hein, L., Reinke, A., Godau, P. et al. (2024). Retrieved from Metrics reloaded: recommendations for image analysis validation: https://www.nature.com/articles/s41592-023-02151-z

Marino, Maria Adele & Avendaño, Daly & Zapata-Julián, Pedro & Riedl, Christopher & Pinker, Katja. (2019). Retrieved from Lymph Node Imaging in Patients with Primary Breast Cancer: Concurrent Diagnostic Tools: https://www.researchgate.net/publication/336533135_Lymph_Node_Imaging_in_Patients_with_Primary_Breast_Cancer_Concurrent_Diagnostic_Tools

Mayo Clinic. (n.d.). Retrieved from Cancer diagnosis and treatment: https://www.mayoclinic.org/diseases-conditions/cancer/diagnosis-treatment/drc-20370594

Mohamed, O. (n.d.). Retrieved from CDD-CESM Dataset. GitHub.: https://github.com/omar-mohamed/CDD-CESM-Dataset?tab=readme-ov-file

Mona Moon Naturals. (2023). Retrieved from A Guide to Check for Bumps, Lumps and Other Changes in Your Breasts: https://www.monamoonnaturals.com/blog/breast-self-exams

Monserrate Intriago-Pazmiño, J. I.-F.-C.-C. (2022). Retrieved from Quantitative Measures for Medical Fundus and Mammography Images Enhancement: https://reunir.unir.net/bitstream/handle/123456789/14338/ijimai8_4_11.pdf?sequence=3

Morin, A. L. (2023). Retrieved from Quantitative imaging: The time is now. Insights into Imaging, 14, Article 103. : https://doi.org/10.1186/s13244-023-01572-w

Moustafa, H. E. (2022). Retrieved from The curated breast cancer Digital Database (CDD-CESM): a collection of contrast-enhanced spectral mammography images for computational imaging research. Scientific Data, 9, Article 416. : https://www.nature.com/articles/s41597-022-01238-0

National Breast Cancer Foundation. (n.d.). Retrieved from Breast cancer facts. : https://www.nationalbreastcancer.org/breast-cancer-facts/

National Cancer Institute. (n.d.). Retrieved from Cancer diagnosis: https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis

Nazmul Ahasan Maruf, A. B. (2025). Retrieved from Breast cancer diagnosis using radiomics-guided DL/ML model-systematic review and meta-analysis: https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2025.1446270/full

Nitesh V. Chawla, K. W. (2002). Retrieved from SMOTE: Synthetic Minority Over-sampling Technique: https://arxiv.org/pdf/1106.1813

Olena Weaver. (2024). Retrieved from Contrast Enhanced Mammography - Course Sample: https://www.youtube.com/watch?v=hcOpOWhBCFE

OpenAI. (2025). Retrieved from https://chat.openai.com/

Philippe Lambin, E. R.-V. (2012). Retrieved from Radiomics: Extracting more information from medical images using advanced feature analysis: https://www.sciencedirect.com/science/article/abs/pii/S0959804911009993

Rizzo, S., Botta, F., Raimondi, S. et al. . (2018). Retrieved from Radiomics: the facts and the challenges of image analysis: https://eurradiolexp.springeropen.com/articles/10.1186/s41747-018-0068-z

Royal Surrey County Hospital. (2018). Retrieved from OPTIMAM1: https://medphys.royalsurrey.nhs.uk/department/optimam/

Siviengphanom, S. &. (2021). Retrieved from Mammography-based Radiomics in Breast Cancer: A Scoping Review of Current Knowledge and Future Needs: https://www.researchgate.net/publication/356264148_Mammography-based_radiomics_in_breast_cancer_a_scoping_review_of_current_knowledge_and_future_needs

Smith, A. J. (2022). Retrieved from Deep learning in cancer imaging: Automated segmentation and diagnosis. Computers in Biology and Medicine, 148, Article 105783. : https://www.sciencedirect.com/science/article/abs/pii/S0010482522007818

The Cancer Imaging Archive. (2021). Retrieved from CDD-CESM | Categorized Digital Database for Low energy and Subtracted Contrast Enhanced Spectral Mammography images: https://www.cancerimagingarchive.net/collection/cdd-cesm/

U.C.Lalij, & Jeukens, C.R.L.P.N. & Houben, Ivo & Nelemans, P.J. & Engen, Ruben & Wylick, E. & Beets-Tan, Regina & Wildberger, J.E. & Paulis, Leonie & Lobbes, Marc. (2015). Retrieved from Evaluation of low-energy contrast-enhanced spectral mammography images by comparing them to full-field digital mammography using EUREF criteria.: https://www.researchgate.net/publication/274714451_Evaluation_of_low-energy_contrast-enhanced_spectral_mammography_images_by_comparing_them_to_full-field_digital_mammography_using_EUREF_criteria

van Griethuysen, J. J. (2017). Retrieved from Computational radiomics system to decode the radiographic phenotype.: https://pubmed.ncbi.nlm.nih.gov/29092951/

van Timmeren, J. E.-L. (2020). Retrieved from Radiomics in medical imaging—“How-to” guide and critical reflection. : https://doi.org/10.1186/s13244-020-00887-2

Wang, Q., Li, K. (2016). Retrieved from Preclinical study of diagnostic performances of contrast-enhaced spectral mammography versus MRI for breast diseases in China.: https://doi.org/10.1186/s40064-016-2385-0

World Cancer Research Fund International. (n.d.). Retrieved from Cancer survival statistics: https://www.wcrf.org/preventing-cancer/cancer-statistics/cancer-survival-statistics/

World Health Organization. (2024). Retrieved from Breast cancer: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

World Health Organization. (2025). Retrieved from Cancer: https://www.who.int/news-room/fact-sheets/detail/cancer

Yu Ji, H. L. (2019). Retrieved from BioMed Central: https://cancerimagingjournal.biomedcentral.com/articles/10.1186/s40644-019-0252-2

Zhang Yongxia, L. F. (2021). Retrieved from Diagnostic Value of Radiomics Analysis in Contrast-Enhanced Spectral Mammography for Identifying Triple-Negative Breast Cancer: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2021.773196/full

Zhao, B., Jiang, L., Xu, Y., & Luo, W. (2023). Retrieved from Radiomics and its feature selection: A review. : https://doi.org/10.3390/sym15101834

Zwanenburg, A. L. (2020). Retrieved from Image biomarker standardisation initiative. Insights into Imaging, 11, Article 91: https://doi.org/10.1186/s13244-020-00887-2