Université Mohamed Boudiaf - M'sila

FACULTE DE TECHNOLOGIE

DEPARTEMENT DE ELECTRONIQUE

Numéro de série……..…

Numéro d'inscription…….……

Thèse

Présentée pour l'obtention du diplôme de

DOCTORAT LMD

Filière : Electronique

Spécialité : Electronique des système embarqués

THEME

Applications of machine learning and deep learning in healthcare: Breast cancer case

Présentée Par

BEGHRICHE Tawfiq

Soutenue le : / /

Devant le jury composé de :

Nom & Prénom	Grade	Etablissement	Qualité
LADJAL Mohamed	**Professeur**	Univ. de **M'sila**	Président
DJERIOUI Mohamed	**MCA**	Univ. de **M'sila**	Encadreur
BRIK Youcef	**MCA**	Univ. de **M'sila**	Co-Encadreur
ASBAI Nassim	**MCA**	Univ. de **Bordj Bou Arreridj**	Examinateur
HAMIDIA Mahfoud	**MCA**	Univ. **Houari Boumedian Babezouar**	Examinateur
OUALI Mohamed Assam	**MCA**	Univ. de **M'sila**	Examinateur
ATTALLAH Bilal	**MCA**	Univ. de **M'sila**	Invité

Année Universitaire : 2024/2025

Acknowledgements

All praise and gratitude are due to ALLAH Almighty for His boundless grace and mercy. It is by His will, strength, and guidance that I was able to persevere through challenges and complete this thesis.

I express my deepest thanks and sincere appreciation to my supervisor **Dr. Mohamed DJERIOUI** and my co-supervisor **Dr. Youcef BRIK** (University of M'sila), whose exceptional mentorship, continuous encouragement, and insightful guidance were instrumental throughout my research journey. Their unwavering support and generosity with their time, despite many responsibilities, had a profound impact on both the direction and quality of this work.

I am thankful to **Pr. Mohamed LADJAL** (University of M'sila) for presiding over the examination committee and for his thoughtful evaluation of this work.

My sincere appreciation also goes to **Dr. Nassim ASBAI** (University of Bordj Bou Arréridj), **Dr. Mahfoud HAMIDIYA** (University of Science and Technology Houari Boumediene, Algiers), and **Dr. Mohamed Assam OUALI** (University of M'sila) for graciously accepting to serve as reviewers. Their time, expertise, and constructive feedback were invaluable to this research.

I also extend heartfelt thanks to the distinguished professors, researchers, and colleagues, in particular, **Pr. Bilal ATTALLAH** (University of M'sila), **Pr. Azzedine ZERGUINE** (King Fahd University, Saudi Arabia), and **Pr. Azeddine BEGHDADI** (Sorbonne Paris Nord University, France), whose insightful contributions were essential in refining my ideas and expanding my perspective.

Lastly, I extend my deepest gratitude to everyone who contributed, directly or indirectly, to the success of this research. Your kindness and support will always be remembered.

May peace and blessings be upon our noble Prophet Muhammad (peace be upon him), his family, and his companions. May Allah bless us all with success and righteousness in this life and the hereafter.

Dedication

I dedicate this work to those who are dearest to me in this world:

To my parents, may they find here the testimony of my deep gratitude and appreciation.

To my brothers, sisters, and family, whose love and support bring joy and strength to my life.

To all my friends who have always encouraged me, and to whom I wish more success.

To all those who have helped me - directly or indirectly - and those who shared with me the emotional moments during the accomplishment of this work and who warmly supported and encouraged throughout my journey.

Thanks!

Tawfiq BEGHRICHE

مُلَخَّص

يُمثِّل سرطان الثدي تحديًا صحيًا عالميًا كبيرًا، ممّا يُبرِز أهمية التشخيص الدقيق والمُبكِّر لتحسين نتائج المرضى. تستكشف هذه الأطروحة تقنيات التعلُّم الآلي (ML) والتعلُّم العميق (DL) المُتقدِّمة لتحسين تشخيص سرطان الثدي، مع التركيز على مهام التصنيف والكشف. تتناول الأطروحة مشاكل تباين البيانات وعدم توازن الأصناف وقابلية التطبيق السريري من خلال تطوير نماذج تشخيصية قوية باستخدام التصوير بالموجات فوق الصوتية ومجموعات البيانات السريرية، مثل مجموعة بيانات سرطان الثدي WBCD ومجموعة بيانات Coimbra. تم تطوير ثلاثة مناهج رئيسية للكشف عن سرطان الثدي: (1) نهج قائم على التعلُّم الآلي يُركِّز على اختيار الميزات وتخفيف عدم توازن الأصناف باستخدام تقنيات مثل SMOTE و KNNOR مع تقييم سبع خوارزميات، بما في ذلك SVM و DNN. (2) طريقة مُتقدِّمة قائمة على التعلُّم العميق تستخدم التعلُّم بالنقل المُعدَّل مع شبكات عصبية التفافية (CNNs) مُدرَّبة مُسبقًا مثل ResNet50 وMobileNetV2 ، تم ضبطها بدقة على مجموعة بيانات صور الموجات فوق الصوتية للثدي (BUSI) لتحسين تصنيف الأورام. (3) نموذج هجين يجمع بين MobileNetV2 و DenseNet121 و InceptionV3 لاستخراج الميزات من صور الموجات فوق الصوتية، تم تنقيحها باستخدام اختيار الميزات القائم على LASSO قبل التصنيف. يتم تقييم أداء النماذج من خلال مقاييس مثل الدقة والحساسية والنوعية والإحكام والاسترجاع وF1-score ، ممّا يُظهِر إمكانات التعلُّم الآلي والتعلُّم العميق في تحسين دقة وأهمية تشخيص سرطان الثدي سريريًا، ممّا يُساعد أطباء الأشعة والأطباء السريريين.

كلمات مفتاحية: سرطان الثدي, التعلّم الآلي, التعلّم العميق, التصوير بالموجات فوق الصوتية, التعلّم بالنقل.

ABSTRACT

Breast Cancer (BC) presents a significant global health challenge, highlighting the importance of timely and accurate diagnosis to improve patient outcomes. This thesis explores advanced machine learning (ML) and deep learning (DL) techniques to enhance breast cancer diagnosis, focusing on classification and detection tasks. It addresses data variability, class imbalance, and clinical applicability by developing robust diagnostic models using ultrasound imaging and clinical datasets, such as the Wisconsin Breast Cancer Dataset (WBCD) and the Coimbra dataset. Three primary frameworks for breast cancer detection are developed: (1) An ML-based approach emphasizing feature selection and class imbalance mitigation using techniques such as SMOTE and KNNOR while evaluating seven algorithms, including Support Vector Machine (SVM) and Deep Neural Networks (DNN). (2) An advanced DL-based method utilizing modified transfer learning with pre-trained convolutional neural networks (CNNs) like ResNet50 and MobileNetV2, reused on the Breast Ultrasound Images (BUSI) dataset for improved tumor classification. (3) A hybrid model combining MobileNetV2, DenseNet121, and InceptionV3 to extract features from ultrasound images, refined using LASSO-based feature

selection before classification. Models performance is evaluated through metrics like accuracy, sensitivity, specificity, precision, recall, and F1-score, demonstrating the potential of ML and DL to enhance the accuracy and clinical relevance of breast cancer diagnostics, aiding radiologists and clinicians.

Key words: Breast Cancer, Machine Learning, Deep Learning, Ultrasound Imaging, Transfer Learning.

RESUME

Le cancer du sein (CS) représente un défi majeur pour la santé mondiale, soulignant l'importance d'un diagnostic rapide et précis afin d'améliorer les résultats pour les patientes. Cette thèse explore les techniques avancées d'apprentissage automatique (AA) et d'apprentissage profond (AP) pour améliorer le diagnostic du cancer du sein, en se concentrant sur les tâches de classification et de détection. Elle aborde les problèmes de variabilité des données, de déséquilibre des classes et d'applicabilité clinique en développant des modèles de diagnostic robustes utilisant l'imagerie ultrasonore et des ensembles de données cliniques, tels que le Wisconsin Breast Cancer Dataset (WBCD) et l'ensemble de données de Coimbra. Trois cadres principaux pour la détection du cancer du sein sont développés : (1) Une approche basée sur l'AA mettant l'accent sur la sélection de caractéristiques et l'atténuation du déséquilibre des classes à l'aide de techniques telles que SMOTE et KNNOR, tout en évaluant sept algorithmes, dont la Machine à Vecteurs de Support (SVM) et les Réseaux Neuronaux Profonds (DNN). (2) Une méthode avancée basée sur l'AP utilisant l'apprentissage par transfert modifié avec des réseaux neuronaux convolutifs (CNN) pré-entraînés tels que ResNet50 et MobileNetV2, affinés sur l'ensemble de données Breast Ultrasound Images (BUSI) pour une meilleure classification des tumeurs. (3) Un modèle hybride combinant MobileNetV2, DenseNet121 et InceptionV3 pour extraire les caractéristiques des images ultrasonores, affinées à l'aide d'une sélection de caractéristiques basée sur LASSO avant la classification. Les performances des modèles sont évaluées à l'aide de mesures telles que la précision, la sensibilité, la spécificité, la justesse, le rappel et le score F1, démontrant le potentiel de l'AA et de l'AP pour améliorer la précision et la pertinence clinique des diagnostics du cancer du sein, aidant ainsi les radiologues et les cliniciens.

Mots clés : Cancer du Sein, Apprentissage Automatique, Apprentissage Profond, Imagerie Ultrasonore, Apprentissage par Transfert.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Full Name
ABC:	Artificial Bee Colony
ACC:	Accuracy
AI:	Artificial Intelligence
ANNs:	Artificial Neural Networks
APTOS:	Aravind Eye Care System
AUC:	Area Under the Curve
AutoML:	Automated Machine Learning
BC:	Breast Cancer
BCDNet:	Breast Cancer Detection Network
BECH:	Breast Cancer Histology
BMI:	Body Mass Index
BANN:	Boosting Artificial Neural Networks
BSense:	Bayesian Hyperparameter Optimized Stacked Ensemble
BreCaHAD:	Breast Cancer Histopathological Annotations for Diagnosis
BreakHis:	Breast Cancer Histopathological Image Classification
BUSI:	Breast Ultrasound Images
CAD:	Computer-Aided Detection/Diagnosis
CBIS-DDSM:	Curated Breast Imaging Subset of the Digital Database for Screening Mammography
CC:	CranioCaudal
CEUS:	Contrast-Enhanced Ultrasound
CLAHE:	Contrast Limited Adaptive Histogram Equalization
CM:	Confusion Matrix
CNNs:	Convolutional Neural Networks
CNV:	Copy Number Variation
Coimbra:	Coimbra Breast Cancer dataset

CSOA:	Crow Search Optimization Algorithm
CT:	Computed Tomography
CWV:	Confidence-Weighted Voting
DBT:	Digital Breast Tomosynthesis
DCNN:	Deep Convolutional Neural Network
DCE:	Dynamic Contrast-Enhanced
DE:	Differential Evolution
DeepMiCa:	Deep Microcalcifications
DL:	Deep Learning
DMR:	Database for Mastology Research
DNN:	Deep Neural Network
DSS:	Decision Support System
DT:	Decision Tree
EBL:	Ensemble Boosting Learning
ELMs:	Extreme Learning Machines
EMS-Net:	Ensemble of MultiScale Convolutional Neural Networks
ERT:	Extremely Randomized Trees
EWT:	Empirical Wavelet Transform
FFPE:	Formalin-Fixed Paraffin-Embedded
FLOPS:	Floating-Point Operations
FN:	False Negative
FNA:	Fine Needle Aspirates
FP:	False Positive
FS:	Feature Selection
GB:	Gradient Boosting
GBDT:	Gradient Boosting Decision Tree
GE:	Gene Expression
GLCM:	Gray-Level Co-occurrence Matrix
GRU:	Gated Recurrent Units
H&E:	Hematoxylin and Eosin
HOMA:	Homeostatic Model Assessment
HPO:	Hyperparameter Optimization
HPFs:	High-Power Fields
HRD:	Homologous Recombination Deficiency

ICPR:	International Conference on Pattern Recognition
IDC:	Invasive Ductal Carcinoma
IoT:	Internet of Things
KCGAN:	Kullback-Leibler Divergence Conditional Generative Adversarial Network
KNN:	k-Nearest Neighbors
KNNOR:	K-Nearest Neighbor Oversampling Ratio
LASSO:	Least Absolute Shrinkage and Selection Operator
LDA:	Linear Discriminant Analysis
LR:	Logistic Regression
LSTM:	Long Short-Term Memory Networks
LWT:	Lifting Wavelet Transform
MC:	Microcalcification
METABRIC:	Molecular Taxonomy of Breast International Consortium
MFO:	Moth-Flame Optimization
MIAS:	Mammographic Image Analysis Society
ML:	Machine Learning
MLP:	Multi-Layer Perceptron
MRI:	Magnetic Resonance Imaging
MTRRE-Net:	Multi-scale Dual Residual Recurrent Network
NAC:	Cancer to Neoadjuvant Chemotherapy
NB:	Naive Bayes
NBIA:	National Biomedical Imaging Archive
PCA:	Principal Component Analysis
Pcam:	PatchCamelyon
PCC:	Pearson Correlation Coefficient
Pre:	Precision
PPV:	Positive Predictive Value
PSCCL:	Penalized Sequential Discriminative Dictionary Learning
PSO:	Particle Swarm Optimization
PTGAN:	Prototype Transfer Generative Adversarial Network
RBF:	Radial Basis Function
RBF-KELM:	Radial Basis Function Kernel Extreme Learning Machine
RBFNN:	Radial Basis Function Neural Network
RBFs:	Radial Basis Functions

RFE:	Recursive Feature Elimination
RF:	Random Forests
RICA:	Reconstruction-Independent Component Analysis
RIDER:	Radiology Image Data from The Cancer Genome Atlas
ROS:	Random Oversampling
ROC:	Receiver Operating Characteristic
RSF:	Random Survival Forest
RUS:	Random Undersampling
RVFL:	Random Vector Functional Link
SAFNet:	Spatial Attention Fusion Network
SEER:	Surveillance, Epidemiology, and End Results
SGD:	Stochastic Gradient Descent
SHAP:	SHapley Additive exPlanations
SiGaAtCNN:	Stacked Generalized Attention Convolutional Neural Network
SNN:	Spiking Neural Network
SMOTE:	Synthetic Minority Oversampling Technique
SOTA:	State-Of-The-Art
SPBC:	Second Primary Breast Cancer
SSF:	Stacked Random Forest
SSL:	Semi-Supervised Learning
ST:	Survival Tree
SVM:	Support Vector Machine
TCGA:	The Cancer Genome Atlas
TCGA-BRCA:	The Cancer Genome Atlas - Breast Invasive Carcinoma
TL:	Transfer Learning
TP:	True Positive
TSBTC:	Thepade's Sorted Block Truncation Coding
UB:	Breast Ultrasound
UCI:	University of California, Irvine
UDIAT:	Unidad de Diagnóstico por Imagen de la Mama
ViTs:	Vision Transformers
WBCD:	Wisconsin Breast Cancer Database
WDBC:	Wisconsin Diagnostic Breast Cancer
WFS:	With Feature Selection

WHO:	World Health Organization
WKNN:	Weighted k-Nearest Neighbor
WOFS:	Without Feature Selection
WPBC:	Wisconsin Prognostic Breast Cancer
WSIs:	Whole-Slide Images
XAI:	Explainable Artificial Intelligence
XGBoost:	eXtreme Gradient Boosting

Chapter 1

INTRODUCTION

1.1 Research Background

Breast Cancer (BC) remains a significant global health challenge, being the second leading cause of cancer-related deaths among women [1]. BC refers to the abnormal, uncontrolled growth of cells in the breast, which can potentially invade surrounding tissues or spread to other parts of the body. This malignancy has been a focal point of medical research due to its widespread occurrence, with more than 2.3 million new cases diagnosed annually, according to the World Health Organization (WHO) [2]. Its high prevalence and mortality rates emphasize the need for refined diagnostic tools to improve early detection, thereby enhancing treatment outcomes [3]. As a result, the medical community continuously seeks ways to improve early detection and treatment methodologies.

Medical imaging plays a crucial role in diagnosing breast cancer, providing detailed insights into internal structures. Various imaging modalities such as mammography, ultrasound, Computed Tomography (CT), and Magnetic Resonance Imaging (MRI) have been central to BC detection. These imaging techniques help visualize breast tissues, distinguishing between healthy and cancerous tissues, and are instrumental in evaluating disease progression and the effectiveness of treatment [4].

Among the most commonly used imaging modalities is mammography, a non-invasive procedure that has been the gold standard in breast cancer screening for decades. Despite its widespread use, mammography presents several challenges, including high rates of false positives and negatives. False positives often lead to unnecessary biopsies and increased patient anxiety, while false negatives can delay diagnosis and treatment. Ultrasound, another prevalent imaging method, is frequently used as a complementary tool to mammography, especially in

dense breast tissues where mammography may be less effective [5]. However, ultrasound's accuracy depends heavily on the operator's expertise, making it a less reliable standalone diagnostic tool. MRI, particularly Dynamic Contrast-Enhanced MRI (DCE-MRI), has emerged as a more advanced technique in evaluating breast cancer, offering superior soft tissue contrast compared to other modalities. DCE-MRI is particularly useful in assessing the response of breast cancer to Neoadjuvant Chemotherapy (NAC), a treatment aimed at shrinking tumors before surgery [6]. This imaging technique allows for the visualization of tumor vascularity and perfusion, providing valuable information about tumor aggressiveness and treatment response. MRI is typically employed both at the initial diagnosis stage and after the administration of NAC to monitor the efficacy of the therapy [7].

Machine Learning (ML) techniques have been notably transformed the field of medical diagnostics, especially in breast cancer classification. These advanced computational techniques facilitate the analysis of extensive datasets, identifying complex patterns and the development of accurate predictions [8]. Deep Learning (DL), a subset of ML, has proven particularly effective in medical imaging due to its capability to automatically extract features from images, thus dismissing the need for manual Feature Selection (FS) [9]. Convolutional Neural Networks (CNNs), a specific type of deep learning model, have demonstrated exceptional success in image classification tasks, making them particularly well-suited for breast cancer classification. CNNs can effectively analyze complex imaging data derived from mammography, ultrasound, and MRI, identifying subtle patterns that may indicate malignancy [10].

Transfer Learning (TL) is a technique that involves transferring the knowledge of a model pre-trained on a different dataset for a specific task—has shown great promise in enhancing the performance of CNNs in breast cancer classification [11]. By leveraging existing knowledge, transfer learning enables models to attain higher accuracy even with smaller datasets, addressing a significant challenge in medical research.

Computer-Aided Diagnosis (CAD) systems have been developed to support radiologists in interpreting mammograms and other imaging modalities. CAD systems employ algorithms to highlight areas of interest within medical images, serving as a supplemental resource to reduce the likelihood of missed diagnoses [12]. Research has demonstrated that combining CAD with traditional diagnostic methods can substantially improve performance rates, particularly in dense breast tissues where mammograms may be less effective [13].

Despite significant advancements, challenges persist in the precise diagnosis of breast cancer. A major limitation is the inherent variability in breast cancer presentations. Tumors can differ widely in size, shape, and location, with their appearance on various imaging modali-

ties influenced by numerous factors, including breast density, tumor type, and the patient's age [14]. Moreover, the imbalance in breast cancer datasets, where negative cases often significantly outnumber positive cases, can adversely affect the performance of machine learning models. To mitigate this issue, techniques such as data balancing were utilized, ensuring that both benign and malignant cases are adequately represented within the dataset to enhance model performance [15].

Additionally, feature selection is essential for developing effective ML techniques for breast cancer classification. By pinpointing the most relevant features from the data, FS reduces model complexity, improves interpretability, and boosts classification accuracy [16]. In the realm of breast cancer, features such as tumor size, shape, texture, and contrast enhancement patterns are vital for differentiating between benign and malignant lesions.

This thesis builds on recent advances in breast cancer classification by examining the integration of deep learning models with advanced imaging techniques. Specifically, the research investigates the application of ML and deep CNNs to classify breast cancer. It includes feature selection, data balancing, and Hyperparameters Optimization (HPO) to improve diagnostic accuracy. The study employs imaging modalities, including ultrasound, in conjunction with clinical data sets to develop and validate predictive models. By merging quantitative imaging data with deep learning approaches, this research aims to provide a more comprehensive evaluation of tumor classification, offering valuable insights for treatment planning and diagnosis. The ultimate objective of this thesis is to contribute to the development of automated breast cancer classification systems that support radiologists and clinicians in making informed decisions. By combining advanced deep learning techniques with state-of-the-art imaging methods, this research aims to enhance the accuracy, efficiency, and reliability of breast cancer diagnostics, eventually improving patient outcomes and reducing the impact of this widespread disease.

1.2 Research Motivation and Significance

Breast cancer remains a significant global health challenge, representing a leading cause of cancer-related mortality among women worldwide. The increasing incidence of this disease underscores the urgent need for improved diagnostic methods that are more accurate, efficient, and less invasive. While conventional diagnostic approaches, such as mammography, ultrasound, and biopsy, play crucial roles in breast cancer classification, they often face limitations in distinguishing benign from malignant lesions with high precision. This can lead to increased rates of false positives, resulting in unnecessary anxiety and invasive procedures, or false nega-

tives, leading to delayed diagnoses and potentially poorer patient outcomes. These limitations highlight the critical need for innovative diagnostic tools that can enhance the accuracy and timeliness of BC classification.

This research is strongly motivated by the transformative potential of machine learning and deep learning in addressing these diagnostic challenges. These advanced computational techniques offer several key advantages. They excel at processing large and complex datasets, identifying subtle patterns and features that may be difficult for human observers to discern. This capability is particularly relevant in medical imaging, where subtle variations in texture, shape, and intensity can be indicative of malignancy. Deep learning models, such as convolutional neural networks, can automatically learn hierarchical representations of image features, eliminating the need for manual feature engineering and potentially achieving superior classification accuracy. Furthermore, breast cancer presents with diverse characteristics across different patients and imaging modalities. DL models can be trained on heterogeneous datasets to learn robust representations that are less sensitive to variations in image quality, patient demographics, and disease stage. This adaptability is crucial for developing diagnostic tools that can generalize well across diverse patient populations. Finally, automated breast cancer classification systems powered by these techniques can significantly reduce the workload of radiologists and clinicians. By providing preliminary assessments and flagging suspicious cases for further review, these systems can streamline diagnostic workflows, improve efficiency, and potentially reduce diagnostic delays.

This research is significant because it aims to develop Artificial Intelligence (AI)-driven frameworks specifically tailored for breast cancer classification using various data modalities (clinical data and ultrasound imaging). This includes addressing specific challenges such as feature selection, class imbalance, and model optimization. In addition, this work aims to contribute to the development of more accurate, efficient and clinically relevant CAD systems for breast cancer. This can lead to accurate classification, faster intervention, and ultimately improved patient outcomes. Finally, this research seeks to advance the state-of-the-art in breast cancer diagnostics by providing a comprehensive evaluation and comparison of different ML and DL techniques. This will contribute to a better understanding of the strengths and limitations of these approaches and inform future research directions. By addressing these key motivations and contributing to these significant advancements, this research aims to make a meaningful impact in the fight against breast cancer.

1.3 Research Objectives

This thesis aims to enhance the accuracy and efficiency of breast cancer classification by developing and evaluating advanced machine and deep learning models. The primary objective is to create robust and clinically relevant computational tools that assist radiologists in making informed decisions. The specific objectives are as follows:

1. Develop an ML-based framework for breast cancer classification using clinical datasets.
2. Deploy a DL-based framework for breast cancer classification using ultrasound images and transfer learning.
3. Design a hybrid framework for efficient breast cancer classification combining feature Extraction (FE), feature selection, and classification.

1.4 Thesis Contributions

This thesis delivers several significant contributions to the field of breast cancer classification by leveraging advanced machine learning and deep learning techniques. These contributions directly address challenges related to data variability, class imbalance, and clinical applicability, ultimately aiming to improve diagnostic accuracy, efficiency, and patient outcomes. The key contributions are as follows:

1. Development of a robust ML-based diagnostic framework that incorporates feature selection based on the correlation of malignancies, data balancing, and Hyperparamters optimization (HPO) for enhancing breast cancer classification using four clinical datasets, including the Wisconsin Breast Cancer Dataset (WBCD), Wisconsin Diagnostic Breast Cancer (WDBC), Wisconsin Prognostic Breast Cancer (WPBC), Coimbra, by:
    - Using a malignancy correlation-based feature selection technique to identify the most relevant predictive features, enhancing model interpretability and performance.
    - Implementing and comparing various class imbalance techniques such as Synthetic Minority Oversampling Technique (SMOTE), K-Nearest Neighbor Oversampling Technique (KNNOR), and Random Undersampling (RUS) to ensure equitable model training and reduce bias.

- Optimizing model parameters through hyperparameters optimization for a range of seven ML algorithms and Deep Neural Networks (DNNs), providing a comprehensive comparative analysis.

2. Development of an advanced DL-based framework for breast cancer classification using modified transfer learning on ultrasound images. This framework addresses the challenges of limited and imbalanced ultrasound datasets, by:

   - Employing modified transfer learning techniques by re-using pre-trained CNN architectures (ResNet50, MobileNetV2, DenseNet121, Xception) on the Breast Ultrasound Images (BUSI) dataset, demonstrating the effectiveness of transfer learning in this context.
   - Providing empirical evidence for the optimal selection and adaptation of pre-trained models for breast ultrasound image analysis.

3. Development of a novel hybrid framework for efficient breast cancer classification combining multi-CNN feature fusion, Least Absolute Shrinkage and Selection Operator (LASSO)-based feature selection, and classification, by:

   - Integrating feature extraction from multiple pre-trained CNNs, such as MobileNetV2, DenseNet121, and InceptionV3, captures diverse image features and fuses them to obtain richer information.
   - Applying LASSO-based feature selection to refine the fused feature vector, leading to improved computational efficiency and potentially enhanced classification accuracy.
   - Demonstrating the efficacy of this hybrid approach in achieving high-precision classification.

4. Comprehensive evaluation and comparative analysis of the proposed frameworks across diverse datasets and using a wide range of evaluation metrics. This contribution provides a thorough assessment of the proposed methodology by:

   - Utilizing a comprehensive suite of metrics, including Accuracy (Acc), Sensitivity (Sens), Specificity (Spec), Precision (Prec), F1-score (F1-s) to provide a holistic view of model performance.
   - Conducting a comparative analysis of the proposed frameworks against each other and potentially against existing state-of-the-art (SOTA) methods.

- Evaluating the generalizability of the models across different data types (clinical and imaging), addressing the issue of data variability.

1.5 Thesis Structure

This thesis is structured into six chapters, each contributing to the overall goal of enhancing breast cancer classification through advanced machine and deep learning techniques. A brief overview of each chapter is provided below:

Chapter 1: Introduction: This chapter provides the context for the research by discussing the global burden of breast cancer and the need for improved diagnostic methods. It outlines the research motivation, objectives, contributions, and the significance of this work. Finally, it presents the structure of the thesis.

Chapter 2: Breast cancer classification: State-of-the-art: This chapter reviews ML and DL techniques in the state-of-the-art applied to breast cancer classification, covering various imaging modalities, relevant preprocessing techniques, traditional ML methods, and advanced DL models. The chapter also identifies existing research gaps and limitations, justifying the proposed research direction.

Chapter 3: Breast cancer classification using feature selection, data balancing and hyperparameters optimization: This chapter presents the development and evaluation of an ML-based framework for breast cancer classification using clinical datasets (WBCD, WDBC, WPBC, Coimbra). The framework incorporates a malignancy correlation-based feature selection technique, class imbalance mitigation strategies (SMOTE, KNNOR, Random Undersampling), and HPO for various ML algorithms, including Support Vector Machines (SVM), K-nearest Neighbors (KNN), Random Forests (RF), Decision Tree (DT), Naïve Bayes (NB), Logistic Regression (LR), and eXtreme Gradient Boosting (XGBoost) along with DNNs.

Chapter 4: Breast cancer classification enhancement using Deep-modified transfer learning: This chapter focuses on the development of a DL-based framework for breast cancer classification using BUSI dataset. The framework leverages modified transfer learning by fine-tuning pre-trained CNN architectures (ResNet50, MobileNetV2, DenseNet121, Xception) to address challenges associated with limited and imbalanced data.

Chapter 5: Deep CNN feature fusion and Lasso-based feature selection for Breast Cancer classification: This chapter introduces a novel hybrid framework for efficient breast cancer classification from ultrasound images. The framework comprises four main stages: feature extraction using multiple CNNs (MobileNetV2, DenseNet121, InceptionV3), feature fusion, LASSO-based

FS, and classification using fully connected layers.

Chapter 6: Conclusion: This chapter summarizes the key findings and contributions of the thesis, discusses the limitations of the proposed approaches, and outlines potential directions for future research.

Chapter 2

BREAST CANCER DETECTION: STATE-OF-THE-ART

2.1 Introduction

Breast cancer remains a leading cause of cancer-related mortality among women worldwide, necessitating continuous advancements in diagnosis and treatment. With the development of machine learning and deep learning techniques, the medical field has seen substantial progress in automating the detection and classification of breast cancer, potentially revolutionizing early diagnosis and treatment planning. This chapter offers a detailed overview of the datasets, pre-processing methods, and classification techniques used in breast cancer research. It starts by detailing the major breast cancer imaging modalities, including mammograms, ultrasound, MRI, and clinical datasets used for research purposes. These datasets are essential for machine learning models, and their variety is crucial for ensuring that diagnostic systems can generalize effectively.

Next, pre-processing techniques, such as data cleaning, normalization, feature selection, feature extraction, and data augmentation, are discussed, focusing on their significance in preparing data for robust classification models. The chapter explores conventional ML methods like SVM, KNN, and RF alongside more advanced deep learning models, including convolutional neural networks, and Long Short-Term Memory (LSTM) networks used for breast cancer classification. This chapter provides a comprehensive literature review, highlighting key advancements and challenges in the field while emphasizing the potential of hybrid approaches to enhance accuracy and robustness in breast cancer classification.

2.2 Breast Cancer Datasets

Breast cancer research has significantly benefited from diverse datasets that enable the exploration of the disease's genetic, clinical, and histopathological dimensions, which helps researchers understand the genetic basis of breast cancer and identify possible biomarkers for diagnosis and treatment [17]. Recent advancements in machine learning have led to developing frameworks that effectively integrate clinical and gene expression data to predict patient outcomes, utilizing various ML algorithms such as decision trees and fuzzy clustering [18]. These datasets and methodologies collectively highlight the evolving landscape of breast cancer research, significantly enhancing diagnostic and predictive capabilities [19].

2.2.1 Clinical Datasets

Clinical breast cancer datasets are essential resources for understanding the complexities of the disease. These datasets often encompass a range of clinical, genomic, and histopathological information, which is crucial for developing and refining diagnostic and therapeutic strategies. Well-known examples include the WDBC and WPBC datasets, which have provided valuable insights for the development of various predictive models [20]. Table 2.1 details several clinical datasets commonly used in recent studies.

2.2.1.1 Wisconsin Diagnostic Breast Cancer

The WDBC is a publicly available dataset in the University of California, Irvine (UCI) machine learning repository [21]. It consists of 569 samples of breast cancer obtained from digitized images of Fine Needle Aspirates (FNA) of breast masses, with 32 attributes described for each sample. The first two columns contain the sample ID and the diagnosis (benign or malignant). The following 30 features are derived from measurements of cell nuclei and include the mean, standard error, and worst values for ten key characteristics, including area, perimeter, radius, texture, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. The dataset comprises 357 benign cases and 212 malignant cases, making it a widely recognized benchmark dataset for research in breast cancer classification [22–44].

2.2.1.2 The Wisconsin Breast Cancer Dataset

The WBCD [45] is a commonly used dataset for breast cancer classification tasks. It contains 699 instances (458 samples were benign, the remaining 241 were malignant), each representing a breast cancer case recorded between 1989 and 1991 at the University of Wisconsin Hospitals.

The dataset consists of 11 features: the first column is a sample ID, while the last column is the class target (benign or malignant). The remaining nine features describe various cell nucleus properties derived from FNA of breast masses, including Uniformity of Cell Shape, Clump Thickness, Bare Nuclei, and Mitoses [22, 23, 25, 27, 29, 31, 33, 36, 38, 42, 46–56].

2.2.1.3 The Wisconsin Prognostic Breast Cancer

The WPBC dataset [57] comprises 198 instances and 34 attributes. Among these are a patient ID, an outcome indicating either "Recur" or "NonRecur", the time to recurrence or health status, and 30 features derived from digitized images of breast masses. These features detail the characteristics of cell nuclei, including area, radius, perimeter, concavity, and texture, comparable to those found in the WDBC dataset. Furthermore, the dataset contains information on tumor size and lymph node status. Tumor size is categorized into four groups (T-1 to T-4) based on diameter, while lymph node status indicates the presence of cancerous nodes. This dataset tracks patients diagnosed with invasive breast cancer who did not have distant metastases, including 151 non-recurrent cases and 47 recurrent cases [23, 27, 28, 30, 31, 33, 38, 53, 58].

2.2.1.4 Coimbra dataset

The Coimbra dataset is a publicly available dataset introduced in 2018 on the UCI machine learning repository [59], commonly used to develop ML models for breast cancer diagnosis. It contains 116 instances with ten attributes derived from blood analysis and anthropometric data. The dataset includes two classes: 52 healthy individuals and 64 breast cancer patients. The features include Body Mass Index (BMI), age, insulin levels, leptin, glucose, adiponectin, MCP-1, resistin, and Homeostatic Model Assessment (HOMA), considered potential indicators for breast cancer [27, 31, 60, 61].

2.2.1.5 Surveillance, Epidemiology, and End Results (SEER)

The SEER dataset [62], offered by the National Cancer Institute's SEER Program, serves as a comprehensive resource for examining breast cancer outcomes and survival trends. It is significant in analyzing cancer survival trends, assessing treatment efficacy, and informing public health strategies. This dataset, sourced from the November 2017 SEER update, includes data on female patients diagnosed with infiltrating ductal and lobular carcinoma between 2006 and 2010. It consists of 4,024 records with 15 attributes, mitigating a wide array of information. Among the recorded cases, 3,408 pertain to living patients, whereas 616 document dead patients. Demographic details encompass variables such as age, race, and marital status, and survival

outcomes. Renowned for its reliability and depth, the SEER breast cancer dataset is an essential tool for researchers and clinicians alike [27, 29, 31, 63].

2.2.1.6 Molecular Taxonomy of Breast International Consortium (METABRIC)

The METABRIC dataset [64] is a comprehensive breast cancer comprising 1,980 patient samples, each with 625 features. These features include Copy Number Alterations (CNA) profiles, gene expression profiles, and clinical data. The dataset categorizes patients into two groups based on their survival outcomes: 491 short-term survivors (labeled as 0) and 1,489 long-term survivors (labeled as 1), with the survival threshold set at five years. The METABRIC dataset is widely used for validating models in breast cancer research for further analysis and experimentation [65–70].

2.2.1.7 Gene Expression (GE)

The GE dataset, as described in [71], is a valuable resource for breast cancer research, offering comprehensive gene expression profiles across 17,814 genes for each sample. The dataset includes 590 samples, comprising 529 breast cancer and 61 noncancer tissue samples. It plays a critical role in cancer biomarker studies, enabling molecular tumor profiling and the identification of gene signatures specific to various breast cancer subtypes. This dataset is instrumental in applying clustering techniques to investigate the heterogeneity of breast cancer, uncovering patterns that distinguish subtypes, advancing personalized treatment strategies, and improving our understanding of breast cancer biology [69, 72–74].

2.2.1.8 RNA-seq

The RNA-seq dataset [75] used for breast cancer research comprises 522 patient samples with 5,567 features, categorizing them based on patient survival outcomes, with 459 short-term survivors labeled as 1 and 63 long-term survivors labeled as 0. The dataset is publicly available on Mendeley for validation and model performance testing [65, 76].

2.2.2 Medical Imaging Datasets

Medical imaging serves as the primary source of data for the detection and diagnosis of cancer, particularly breast cancer. The accurate and precise identification of BC is crucial for timely detection, significantly improving the chances of survival. Numerous imaging techniques have been created and continually improved to detect breast cancer in its early stages. The main

Table 2.1: Comparison of clinical datasets for breast cancer diagnosis within recent studies.

Datasets	**References**	**Classes**	**Samples**	**Samples per Class**	**Features**	**Advantages**	**Limitations**
WBCD	[22, 23, 25, 27, 29,31,33,36,38, 42, 46–56]	2	699	B (458), M (241)	10	Widely used benchmark dataset for consistent model evaluation.	Limited to clinical features; lacks imaging and genomic data, restricting multimodal analysis.
WDBC	[22–44]	2	569	B (357), M (212)	32	Standardized dataset with well-annotated clinical features.	Exclusively clinical data without imaging, limiting comprehensive diagnostic insights.
WPBC	[23, 27, 28, 30, 31,33,38,53,58]	2	198	B (151), M (47)	34	Includes follow-up data for survival and recurrence studies.	Small sample size, lacks imaging and genomic information.
Coimbra	[27, 31, 60, 61]	2	116	H (52), NH (64)	10	Includes metabolic and clinical data, valuable for breast cancer risk assessment.	Small dataset without multimodal validation limits broader applicability.
SEER	[27, 29, 31, 63]	2	4024	A (3408), D (616)	15	Extensive demographic and clinical attributes for survival analysis.	No imaging or genomic data, limiting integration with advanced diagnostic models.
GE	[69, 72–74]	2	Variable	Variable	Variable	Enables early detection, personalized treatments through molecular data.	Limited clinical context; lacks multimodal validation for generalizability.
METABRIC	[65–70]	Variable	Variable	Variable	Variable	Large-scale genomic data enhances personalized treatment strategies.	Primarily gene expression data without imaging integration.
RNA-seq	[65, 76]	Variable	Variable	Variable	Variable	Facilitates molecular profiling for personalized diagnosis.	Limited population generalizability; lacks multimodal data validation.

imaging methods currently employed for BC diagnosis include mammography, ultrasound, MRI, histopathology, and thermography, as illustrated in Figure 2.1.

2.2.2.1 Mammograms

Mammography is widely recognized as the most common and effective method for early detection of breast cancer in asymptomatic women [19]. This technique utilizes low-energy X-rays to examine and screen breast tissue [77], effectively identifying microcalcifications and clusters of calcifications, which are often indicative of breast cancer. Several publicly available mammography datasets are frequently used in research. Table 2.2 provides a detailed analysis of these datasets.

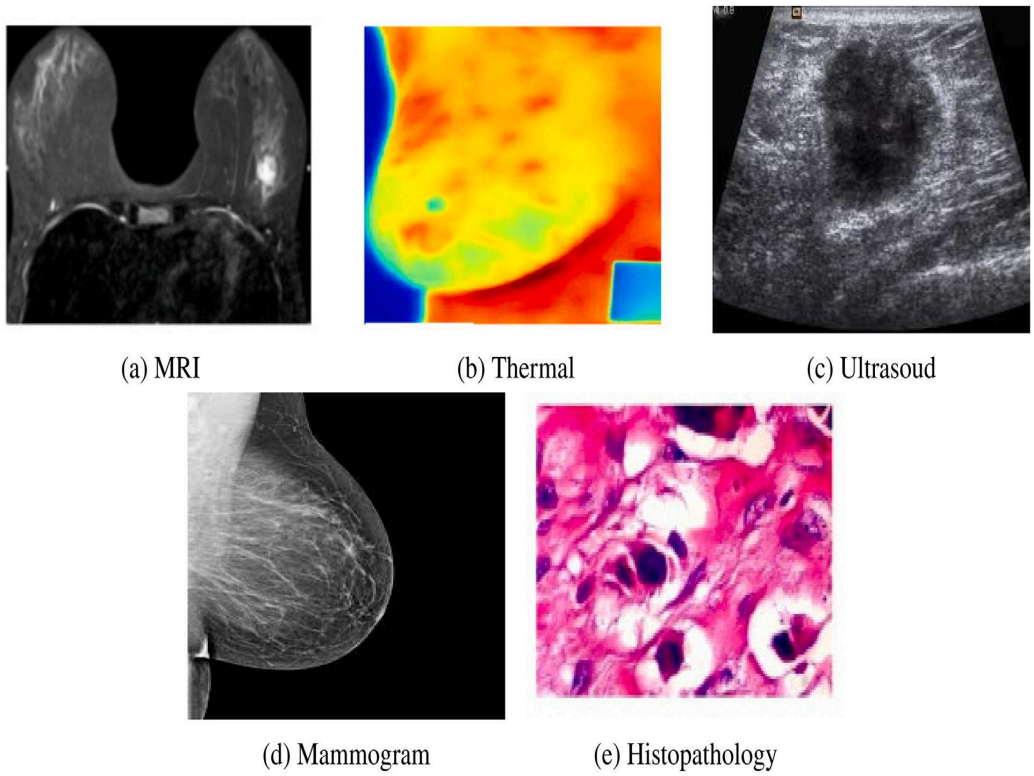


Figure 2.1: Breast cancer images obtained by different imaging techniques: (a) MRI (b) thermography (c) ultrasound (d) mammograms (e) histopathology.

Digital Database for Screening Mammography (DDSM): The DDSM, provided by the University of South Florida [78], is a key public dataset for breast cancer research, featuring over 10,000 digitized screen-film mammograms. It includes 2,500 normal, 3,000 benign, and 4,500 malignant images captured from multiple views (MLO and CC) at a resolution of 3000 $\times$ 4800 pixels. Each case contains detailed annotations from radiologists, including BI-RADS scores, breast density, lesion types, and localization of abnormalities. DDSM is essential for tumor detection and classification, particularly in developing computer-aided detection (CAD) systems for identifying breast cancer in dense breast tissue [79–91].

Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM): The CBIS-DDSM [92] dataset, curated by the University of South Florida, is an updated version of the DDSM. It comprises 3,103 mammographic images from 1,566 patients, focusing on masses and calcifications. The dataset includes pixel-level ROI annotations for benign and malignant cases, DICOM format images, ACR density labels, biopsy-proven pathology results, and BI-RADS scores. It features mediolateral oblique (MLO) and craniocaudal (CC) views for each breast. It is a vital resource for developing and bench-

marking breast cancer detection models, particularly in segmentation and classification tasks [83, 93–97, 97–101].

INbreast: The INbreast [102] dataset is a comprehensive collection of 410 full-field digital mammograms from 115 patients obtained by the Centro Hospitalar de S. João, Porto, Portugal. It contains normal and abnormal cases, with detailed annotations for masses, calcifications, asymmetries, and distortions. Each abnormal case is precisely labeled with ground truth annotations, specifying lesion type, location, and BI-RADS score. The dataset provides high-quality images with MLO and CC views, making it a valuable resource for breast cancer detection and classification research [80, 81, 83, 90, 93–95, 97, 103–107].

Mammographic Image Analysis Society (MIAS): The MIAS dataset [108] is a widely utilized resource for research in breast cancer diagnosis. It includes 322 digitized mammograms from 161 patients, classified into normal (207), benign (63), and malignant (52) cases. The images are provided in PGM format and have a resolution of 1024 × 1024 pixels. The dataset incorporates radiologist annotations indicating abnormalities such as masses, calcifications, and asymmetries, along with tissue density labels: fatty, fatty-glandular, and dense-glandular [23, 47, 80, 84, 86, 87, 90, 91, 94, 103, 105, 106, 109–111].

Table 2.2: Comparison of mammogram datasets for breast cancer diagnosis within recent studies.

Datasets	**References**	**N. of Classes**	**N. of Images**	**Images per Class**	**Advantages**	**Limitations**
MIAS	[23,47,80,84,86,87, 90, 91, 94, 103, 105, 106, 109–111]	3	322	N (207), B (63), M (52)	Public, good for initial testing; labeled abnormalities aid early training.	Small, limited diversity; low resolution limits its advanced applications.
DDSM	[79–91]	3	1500	N (519), B (479), M (502)	Large, diverse, valuable for robust feature extraction and diagnostic modeling.	Requires significant preprocessing; dataset age may affect image quality.
INBreast	[80, 81, 83, 90, 93–95, 97, 103–107]	3	410	B (220), N (67), M (49)	High-resolution images, excellent for deep learning and detailed analysis.	Limited access/high cost; small size limits generalizability.
CBIS-DDSM	[83, 93–97, 97–101]	3	2,620	B (870), N (695), M (914)	Comprehensive, detailed annotations, ideal for deep/federated/transfer learning.	Requires complex preprocessing; high variability may challenge model consistency.

2.2.2.2 Ultrasound

Ultrasound imaging serves as a valuable auxiliary tool to mammography in breast cancer detection. Unlike mammograms, ultrasound images are typically grayscale and may exhibit less detail. However, ultrasound excels at differentiating between cystic and solid masses, a capability where mammography can be less effective. It is particularly advantageous for assessing dense breast tissue, where mammography's sensitivity can be limited [112]. Several ultrasound datasets are commonly used in research and described in Table 2.3.

Breast Ultrasound Images: The BUSI dataset, gathered in 2018 at Baheya Hospital in Cairo, Egypt [113], plays a crucial role in research focused on breast cancer detection. It comprises 780 ultrasound images from women aged 25 to 75, all captured using LOGIQ E9 systems. The dataset includes ground truth masks in PNG format, with an approximate resolution of 500 $\times$ 500 pixels. It contains 437, 210, and 133 images labeled as benign, malignant, and normal, respectively, making it an invaluable resource for studies aimed at differentiating between tumor types while employing safe, radiation-free ultrasound imaging [88, 114–131].

UDIAT: The UDIAT dataset [132] is a collection of breast ultrasound images utilized for research in breast cancer detection, containing 163 pictures, each corresponding to a single lesion accompanied by a ground truth mask, providing detailed annotations of the lesions, categorized into 109 benign and 54 malignant cases. These images were acquired using a Siemens ACUSON scanner at the UDIAT Diagnostic Centre of the Parc Tauli Corporation in Sabadell, Spain. This dataset has been instrumental in developing and evaluating machine learning models for breast cancer detection, particularly in segmenting breast masses in ultrasound images. The dataset's detailed annotations and high-quality photos make it a valuable resource for advancing automated breast cancer detection methods [122].

Breast Ultrasound Video Dataset: Breast US videos were collected for a retrospective study at Shengjing Hospital between March 2021 and May 2022, involving 332 patients with confirmed breast lesions. The study received ethics committee approval, and all participants provided informed consent for Contrast-Enhanced Ultrasound (CEUS) and ultrasound-guided biopsy [133]. The dataset includes 1,000 video sequences from Linyi People's Hospital (LPH) and the Second Affiliated Hospital of Harbin Medical University (SAHHMU), showcasing transverse and longitudinal views with an average duration of 7 seconds. Lesions were imaged using a Logiq E9 Doppler ultrasound machine before surgery and biopsy, facilitating precise diagnoses confirmed by coarse needle biopsy [134].

Table 2.3: Comparison of ultrasound datasets for breast cancer diagnosis within recent studies.

Datasets	References	N. of Classes	N. of Images	Images per Class	Advantages	Limitations
BUSI	[88, 114–131]	3	780	B (437), M (210), N (133)	Public, labeled (B, M, and N); suitable for multi-class classification/segmentation.	Small size/resolution, limited real-world applicability/diversity; minimal pathological annotations.
Breast US Video	[133, 134]	3	332	B (101), M (102), MS(129)	Provides temporal data for real-time training; includes diverse cases.	Small volume, high computational cost for video processing.
UDIAT	[122]	2	163	B (110), M (53)	High-quality, verified annotations; suitable for binary classification.	Small size/diversity limits generalizability; lacks video data.

2.2.2.3 Magnetic Resonance Imaging

MRI is a non-invasive imaging technique that uses a magnetic field and radio waves to produce detailed images of the breast [135]. This modality is particularly useful for screening dense breast tissue and is often used with mammography and ultrasound to improve diagnostic accuracy [136]. Due to its high sensitivity, MRI can detect cancerous tumors that mammography may miss, making it a crucial screening tool for high-risk individuals. Table 2.4 presents the most commonly used MRI datasets.

Dynamic Contrast-Enhanced MRI: The DCE-MRI dataset [137], provided by the Cancer Imaging Archive, is a publicly accessible collection of breast imaging acquired using a 1.5-T Signa scanner with a bilateral phased-array breast mass, obtaining the dataset's features as high-quality images through a dynamic contrast-enhanced protocol that combines low temporal and high spatial resolution T1-weighted sequences. These images were taken at three-time points: T0 (pre-contrast), T1, and T2 (post-contrast), enabling analysis of contrast dynamics at 0, 2.5, and 7.5 minutes, making the DCE-MRI dataset a valuable resource for enhancing breast cancer research, especially in lesion detection, segmentation, and pre-surgical planning [76, 138, 139].

RIDER: The RIDER [140] dataset is a collection of MRI scans, including tumor and healthy images, acquired from the National Biomedical Imaging Archive (NBIA), managed by the United States National Cancer Institute (USNCI). The original dataset contains 1,500 images from five patients, with their corresponding ground truth annotations. For research purposes, 90 tumor images and their annotated counterparts are commonly used in experiments to develop and evaluate medical imaging models.

The Cancer Genome Atlas (TCGA): The TCGA dataset [141] is a large-scale publicly available database, including various molecular and clinical data for multiple cancer types. It has several sub-datasets, as follows:

TCGA-BRCA is a dataset focused on breast cancer, containing gene expression data, Copy Number Alterations (CNAs), and clinical information for over 1,000 patients. It is commonly used for molecular subtype classification and survival analysis [67, 68, 142].

TCGA-CRCk centers on colorectal cancer, providing Microsatellite Instability (MSI) status, histopathological images, and genomic data. It is frequently used for genomic label classification and machine learning tasks [143].

TCGA-BC refers to a breast cancer sub-dataset with Formalin-Fixed Paraffin-Embedded (FFPE) and Whole-Slide Images (WSIs) used to study Homologous Recombination Deficiency (HRD), genomic features, and tumor heterogeneity. This dataset includes genomic data like HRD scores and is often employed in studies related to personalized therapy predictions [143].

Table 2.4: Comparison of MRI datasets for breast cancer diagnosis within recent studies.

Datasets	**References**	**N. of Classes**	**N. of Images**	**Images per Class**	**Advantages**	**Limitations**
TCGA-BRCA	[65, 67, 68, 142]	4	Variable	Variable	Public, extensive multi-omics data enables multi-modal models.	Requires extensive preprocessing/computation; limited imaging/genetic diversity, inconsistent data collection.
RIDER	[96]	2	1,500	H (-), NH (-)	Multi-timepoint DCE-MRI data for studying cancer progression.	Relatively small, potentially limited generalizability, inconsistent acquisition protocols.
DCE-MRI	[76, 138, 139]	4	Variable	Variable	Rich temporal contrast aids tumor detection/characterization/treatment planning.	High acquisition cost, limited size, resource-intensive processing.
TCGA-BC	[143]	Variable	Variable	Variable	Provides breast/colorectal cancer data for multi-cancer analysis.	Limited imaging diversity; combining cancer types may introduce noise.

2.2.2.4 Histopathology

Histopathology is considered the gold standard imaging modality for breast cancer diagnosis, providing crucial phenotypic information for treatment planning [144]. However, it presents challenges for multi-class classification due to high similarity among cancerous cells, significant

intra-class variability, and subtle distinctions between classes. Images within the same class often exhibit higher resolution, contrast, and significant appearance variations compared to images from different classes, complicating differentiation. Furthermore, gigapixel WSIs, which can exceed 1 GB in size, pose computational challenges for deep learning models when attempting thorough processing [9]. Table 2.5 illustrates commonly used histopathology datasets.

Breast Cancer Histopathological Image Classification (BreakHis): The BreakHis dataset, introduced by Spanhol et al. [145], is a publicly accessible collection of 7,909 high-resolution histopathology images of breast cancer obtained from 82 patients. It consists of 2,480 benign and 5,429 malignant tumor images, and both categories are further divided into four distinct subtypes, enabling detailed subtype-specific analysis. These images were constructed using an Olympus BX-50 microscope at four magnification levels: 40x, 100x, 200x, and 400x, stored in PNG format at a resolution of 700 × 460 pixels in RGB color, ensuring compatibility with a wide range of computational models and tools. With its diverse collection of well-annotated images, the BreakHis dataset has become a cornerstone resource particularly valuable for advancing research in breast cancer pathology and computer-aided diagnosis [146–174].

Breast Cancer Histopathological Annotations for Diagnosis (BreCaHAD): This dataset was introduced by Nazeri et al. [175], includes 1,662 Hematoxylin and Eosin (H&E) stained histopathology images of breast tissue from 82 patients with Invasive Ductal Carcinoma (IDC) at Shahid Beheshti University. It includes low-resolution (×4) and high-resolution (×40) images, totaling 8,310 image patches. High-resolution images come with detailed annotations on tumor cells, stroma, and lymphocytes, aiding research in tumor segmentation, cell detection, and breast cancer grading [151, 174].

Camelyon16: The Camelyon16 dataset [176], introduced by Bejnordi et al. as part of the CAMELYON16 challenge, obtaining 400 whole-slide images of B-lymph node sections stained with hematoxylin and eosin. It includes 270 WSIs for training and 130 for testing, sourced from Radboud UMC and UMC Utrecht. The dataset offers pixel-level annotations for metastatic regions, serving as a benchmark for algorithms that detect lymph node metastases in breast cancer patients with slides captured at 40x magnification [177–179].

PatchCamelyon (Pcam): The PCAM dataset [180], developed by Veeling et al., consists of 327,680 color images sourced from the Camelyon16 challenge, which features 400 whole-slide images of H&E stained lymph node sections. Each image in the dataset measures 96 × 96 pixels

and is organized into training, validation, and test subsets, ensuring a balanced distribution of positive and negative labels. Positive labels indicate the presence of tumor tissue within a central 32 × 32-pixel area, while negative labels denote its absence. This labeling method supports convolutional models in their tasks related to breast cancer detection [181, 182].

MITOS-ATYPIA-14: The MITOS-ATYPIA-14 dataset [183], released by the International Conference on Pattern Recognition (ICPR), was designed for challenges in breast cancer detection. The dataset consists of H&E stained slides scanned using two slide scanners: Hamamatsu Nanozoomer 2.0-HT and Aperio Scanscope XT, at different magnifications of 10x, 20x, and 40x. It includes 1,696 High-Power Fields (HPFs) at 40x magnification, with a resolution of 1539x1376 pixels. The training set contains 1,200 HPFs with 749 labeled mitotic cells, annotated by two pathologists, with a third pathologist consulted in case of disagreement. This dataset is commonly used in breast cancer detection research [181, 182, 184, 185].

Breast Cancer Histology (BACH): The BACH dataset [186], developed for the ICIAR 2018 grand challenge, consists of H&E-stained breast histology images, including 400 training images (2048x1536 pixels) acquired with a DM 2000 LED microscope and Leica ICC50HD camera from three hospitals in Portugal, labeled by two medical experts, categorized into four classes: Normal, Benign, In situ carcinoma, and Invasive carcinoma. Thus, it is widely used to enhance breast cancer classification and diagnosis [155, 159, 165, 166].

Invasive Ductal Carcinoma (IDC): The IDC dataset [187] is a publicly available dataset that contains 162 whole-slide images from breast cancer specimens, resulting in 277,524 image patches, divided into two distinct categories: IDC-negative (198,738 patches) and IDC-positive (78,786 patches), all digitized at a magnification of 40x with the size of 50x50 pixels. It features reliable annotations from medical specialists, making it commonly used for developing automatic detection algorithms for IDC in breast cancer pathology [165, 188–190].

2.2.2.5 Thermography

Breast thermal imaging, or thermography, is another imaging modality used for breast cancer detection [191]. This technique detects thermal abnormalities, indicating potential breast abnormalities due to increased heat production by cancer cells. Unlike other breast cancer imaging techniques, thermography is non-invasive, painless, and safe, contributing to early detection efforts [192]. Table 2.6 presents the most commonly used thermography datasets.

Database for Mastology Research (DMR): The DMR is a publicly accessible platform dedicated to breast cancer research, specifically designed to store hematologic images for initial identification purposes. The dataset [193] comprises both thermal and mammography images, with a focus on thermograms categorized into healthy (745 images) and sick (261 images) subjects, all at a resolution of 640 × 480 pixels.

DMR with Infrared Images (DMR-IR): The DMR-IR is a specialized subset of DMR focused on thermogram images for breast cancer detection. It contains 1,345 images collected from women aged 32-74 and categorized into three groups: 705 normal images, 200 benign tumor images, and 440 malignant tumor images. All thermal infrared images, with a resolution of 640 × 480 pixels, have been verified by radiologists. For research purposes, 1,000 images are allocated for training, while the remaining 345 images are devoted to testing and validation phases [194, 195].

Table 2.5: Comparison of histopathology datasets for breast cancer diagnosis within recent studies.

Datasets	**References**	**N. of Classes**	**N. of Images**	**Images per Class**	**Advantages**	**Limitations**
BreakHis	[146–174]	2	7,909	B (2,480), M (5,429)	High-resolution, good for detailed tumor analysis, multi-task learning.	Limited to histopathology, lacks multi-modal data/demographic diversity.
BreCaHAD	[151, 174]	2	162	M (113), B (49)	Enhances robustness across tumor subtypes, useful for general cancer detection training.	Small size, potential data imbalance, limited diversity.
BACH2018	[155, 159, 165, 166]	4	400	N, B, INC, INV	Supports multi-class classification for improved diagnostic generalization.	Limited to histopathology, not applicable to other modalities.
MITOS-ATYPIA-14	[181, 182, 184, 185]	2	1,000	Variable	Specializes in mitosis detection for assessing cancer aggressiveness.	Small size limits generalization.
CAMELYON16	[177–179]	2	400	MT, NMT	High-quality lymph node metastasis annotations, ideal for segmentation.	Computationally intensive due to large whole-slide images.
PCAM	[181, 182]	2	327,680	MT, NMT	Extensive annotations, allows generalization across tumor features.	Very large size, potential class imbalance.
IDC	[165, 188–190]	2	277,524 patches	IDC+ (78,786), IDC- (198,738)	Improves robustness for IDC subtypes, good for real-world clinical data training.	Patch imbalance may affect performance, especially for IDC subtypes.
Mitosis/Non-mitosis	[196]	2	150	MI (75), NMI (75)	Focused on mitosis detection, aiding in aggressive cancer identification.	Small size limits generalization.

Table 2.6: Comparison of thermography datasets for breast cancer diagnosis within recent studies.

Datasets	**References**	**N. of Classes**	**N. of Images**	**Images per Class**	**Advantages**	**Limitations**
DMR	[195]	2	1006	H (745), NH (261)	Supports binary classification; captures temperature variations.	Limited sick cases, thermal-only imaging reduces robustness/generalizability.
DMR-IR	[194]	3	1345	N (705), B (200), M (440)	Facilitates multi-class classification with a relatively large sample size; includes malignant cases.	Class imbalance, reliance on thermal imaging may limit diagnostic accuracy.

2.3 Preprocessing

Pre-processing is crucial in preparing data for effective machine learning models, particularly in medical applications like breast cancer diagnosis. This section explores various pre-processing techniques applied to clinical and imaging datasets. These techniques enhance data quality, reduce noise, handle missing values, and extract relevant features to improve model performance and diagnostic accuracy. Common methods include data cleaning, normalization, feature selection and extraction, and data augmentation, each tailored to the specific characteristics of the data.

2.3.1 Data Normalization

Data normalization is a crucial preprocessing step that transforms feature values to a consistent scale, preventing features with larger ranges from dominating the learning process and ensuring equitable contributions from all features, which is particularly important for distance-based algorithms and high-dimensional datasets like microarray and RNA-seq data, where normalization mitigates biases and enables accurate comparisons. Various normalization methods exist, each with distinct effects like the standard techniques like Min-Max scaling, which scales features to a 0–1 range, and Z-score normalization, which standardizes data to a mean of 0 and a standard deviation of 1, are widely employed to improve model stability and performance [197]. As shown in Table 2.7, these methods are applied across both clinical and imaging data, though some are more prevalent in one domain or the other. While these methods improve model performance, they handle outliers differently. Adaptive techniques, which dynamically adjust normalization parameters based on the data, have emerged as promising solutions for addressing

outlier sensitivity and enhancing performance in complex decision-making scenarios [198].

Table 2.7: Summary of normalization techniques and their effects on clinical and imaging data.

Preprocessing Technique	**References**	**Clinical**	**Imaging**	**Effect**
Min-Max Normalization	[23,31,33,34,39,40, 58, 67, 70, 74, 199]	Yes	No	Scales features to 0–1, improving model stability and performance for algorithms sensitive to feature scales.
	[88, 97, 119, 120, 122, 189, 190, 200–202]	No	Yes	
	[76]	Yes	Yes	
z-score Normalization	[24, 32, 39, 43, 63, 107, 199]	Yes	No	Standardizes data to mean 0 and standard deviation 1, enhancing performance and mitigating outliers.
	[94, 111, 142, 149, 168]	No	Yes	
	[65]	Yes	Yes	
Macenko	[157, 171, 181]	No	Yes	Stain normalization for histopathology images, reducing staining variability and improving feature consistency.
Global Contrast Normalization (GCN)	[94, 153]	No	Yes	Enhances image contrast robustness by normalizing pixel intensity using mean and standard deviation.
Correlation Coefficient-based Normalization	[52]	Yes	No	Reduces multicollinearity by normalizing features based on their correlation with targets or other features.
Arithmetic Mean Scaling Technique	[40]	Yes	No	Scales features relative to their mean, emphasizing feature proportions.
Equilibration Scaling Technique	[40]	Yes	No	Adjusts feature distribution for improved uniformity, potentially enhancing algorithm performance.
Geometric Mean Scaling Technique	[40]	Yes	No	Similar to arithmetic mean scaling but less sensitive to outliers.
The LIMMA (Linear Models for Microarray Data)	[73]	Yes	No	Normalizes gene expression data, managing experimental design and small sample sizes effectively.
Reinhard	[181]	No	Yes	Transfers color properties in histopathology images, reducing color variation between slides.
LAHE	[159]	No	Yes	Enhances local image contrast, useful for varying illumination or contrast.
Retinex	[159]	No	Yes	Enhances contrast and color constancy by estimating reflectance independent of illumination.
Vahadane	[203]	No	Yes	Reduces stain variability in histopathology images via color deconvolution.
Fuzzy-based KNN	[72]	Yes	No	Uses fuzzy logic to improve KNN's robustness to noise and outliers in gene expression data.
Log-transformation	[63]	Yes	No	Compresses skewed data, stabilizes variance, and suits statistical models.
Feature-wise Standardization	[119]	No	Yes	Equivalent to z-score normalization, applied feature-by-feature.
Median Absolute Deviation (MAD) Normalization	[46]	Yes	No	Robust against outliers, normalizes using the median and MAD.

2.3.2 Data Cleaning

Data cleaning, or data cleansing, is a critical preprocessing stage focused on identifying and rectifying errors, inconsistencies, and inaccuracies within datasets. This process is essential for ensuring data quality and reliability, which are fundamental for accurate analysis and informed decision-making. Data cleaning can consume a significant portion of the data analysis workflow, often up to 80%, underscoring the importance of robust and efficient cleaning methodologies [204]. Table 2.8 summarizes various data cleaning techniques used in breast cancer studies.

Table 2.8: Summary of data cleaning techniques and their effects on clinical and imaging data.

Preprocessing Technique	References	Clinical	Imaging	Effect
Handle missing values (Removal)	[25, 27, 29, 31, 32, 39, 43, 63, 205]	Yes	No	Reduces dataset size by removing instances with missing data, potentially simplifying analysis but risking information loss.
Handle missing values (Not specified)	[33, 44, 49, 60, 66, 199]	Yes	No	Essential for preparing datasets; effects depend on method used (imputation or removal).
Outliers removal	[43, 44, 51, 55, 60]	Yes	No	Improves performance by removing extreme values but risks discarding valuable data.
Handle missing values (Imputation: median value)	[29, 52, 206]	Yes	No	Simple and robust to outliers; reduces variance and preserves dataset size.
Handle missing values (kNNImputer)	[28, 65]	Yes	No	Uses k-nearest neighbors to impute missing values; captures feature relationships but computationally intensive.
Handle missing values (weighted nearest neighbor)	[67, 70]	Yes	No	Weighted variant of kNN for more accurate imputation.
Handle missing values (average/most frequent)	[31]	Yes	No	Simple methods for numerical/categorical features; can introduce bias.
Handle missing values (missForest)	[63]	Yes	No	Non-parametric, accurate imputation using random forests; computationally intensive.
Handle missing values (simple linear regression)	[58]	Yes	No	Predicts missing values based on linear relationships; limited to linear correlations.
Categorical-numerical conversion (one-hot encoding)	[63, 206]	Yes	No	Converts categorical variables into binary columns; increases dimensionality.
Unlabeled data removal	[111]	No	Yes	Removes unlabeled data to ensure reliable training; reduces dataset size.
Removing invalid mammograms/masks	[98]	No	Yes	Ensures accuracy of mammograms and masks by removing invalid pairs.
Quality control (QC)	[73]	Yes	No	Removes unreliable rows; improves quality at the cost of dataset size.
Not specified	[68]	Yes	Yes	Not specified.

2.3.3 Feature Selection

Feature selection is an important preprocessing step in machine learning that aims to enhance model performance by identifying the most relevant features from a dataset. By reducing

dimensionality, FS mitigates overfitting, decreases computational costs, and improves model generalization and interpretability. FS methods are broadly categorized into filter, wrapper, and embedded approaches [207, 208]. The benefits of FS extend to noise reduction, model simplification, and enhanced interpretability, making it particularly valuable in domains like bioinformatics and healthcare, where understanding feature importance is paramount [209].

Filter methods assess feature relevance based on inherent data characteristics, offering computational efficiency. Wrapper methods employ ML algorithms to evaluate feature subsets, often achieving higher accuracy but at a greater computational cost. Embedded methods integrate FS directly into the model training process, striking a balance between accuracy and efficiency. A variety of FS techniques, along with their application to clinical and/or imaging data and their respective effects, are summarized in Tables 2.9 and 2.10.

Table 2.9: Summary of feature selection techniques and their effects on clinical and imaging data (Part 1).

Preprocessing Technique	**References**	**Clinical**	**Imaging**	**Effect**
Correlation-Based Feature Selection	[32, 51, 52, 58, 60, 65, 205]	Yes	No	Reduces redundancy by selecting highly correlated features with the target, improving interpretability and reducing overfitting.
	[65]	Yes	Yes	
Principal Component Analysis (PCA)	[142]	No	Yes	Transforms features into principal components ranked by variance, reducing dimensionality and enhancing efficiency.
	[31, 48, 52]	Yes	No	
PCA + Linear Discriminant Analysis (LDA)	[80, 110]	No	Yes	Combines PCA for dimensionality reduction and LDA for class separability, improving classification performance.
LASSO Logistic Regression	[72]	Yes	No	Shrinks coefficients of less important features to zero, creating a sparse and interpretable model.
	[210]	No	Yes	
Minimum Redundancy Maximum Relevance (mRMR)	[67]	Yes	Yes	Selects features with high relevance and low redundancy for a concise and informative set.
	[70]	Yes	No	
Polynomial Kernel PCA (KPCA)	[69]	Yes	No	Captures non-linear relationships in data, improving performance on complex datasets.
	[87]	No	Yes	
SULOV Algorithm	[61]	Yes	No	Selects features with high variance and minimal redundancy, enhancing efficiency and performance.
Information Gain Directed Simulated Annealing Genetic Algorithm Wrapper (IGSAGAW)	[22]	Yes	No	Combines information gain with optimization algorithms for improved feature subset selection.
Weighted Heterogeneous Value Distance Metric (WHVDM)	[211]	No	Yes	Calculates instance distances using a weighted metric, enhancing feature and instance selection.

Table 2.10: Summary of feature selection techniques and their effects on clinical and imaging data (Part 2).

Preprocessing Technique	**References**	**Clinical**	**Imaging**	**Effect**
Logistic Regression	[23]	Yes	No	Determines feature importance based on coefficient magnitude, suitable for linear relationships.
Neural Network	[24]	Yes	No	Captures complex non-linear feature relationships through connection weights.
Variable Importance Measure (VIM)	[25]	Yes	No	Quantifies feature importance in tree-based models by contribution to impurity reduction.
Optimization Algorithms (ESO, GSO, PSO)	[26, 199]	Yes	No	Metaheuristic methods inspired by nature for feature subset optimization.
Chi-square	[60]	Yes	No	Selects features significantly associated with the target, suitable for categorical data.
Linear Regression (LR)	[72]	Yes	No	Assesses feature importance using coefficients; suitable for linear relationships.
Random Forest	[212]	No	Yes	Measures feature importance by impurity reduction, providing robust results.
Gorilla Troops Optimizer (GTO)	[33]	Yes	No	Metaheuristic optimization algorithm inspired by gorilla behavior, used as a wrapper for feature selection.
Mutual Information	[51]	Yes	No	Identifies features with high statistical dependence on the target variable, capturing both linear and non-linear relationships.
Gradient Boosting Decision Tree-Based Mayfly Optimization (GBDTMO)	[36]	Yes	No	Integrates GBDT and optimization algorithms for tailored feature selection.
Recursive Feature Selection Algorithm (RFE)	[39]	Yes	No	Iteratively removes the least important features, refining the selection for performance.
Differential Gene Expression Analysis (DGEA)	[73]	Yes	No	Identifies significant gene associations using statistical tests.
Whale Optimization Algorithm (WOA), Dragonfly Algorithm (DA)	[42]	Yes	No	Metaheuristic algorithms for efficient feature subset selection.
Stacked Feature Selection (SFS)	[68]	Yes	Yes	Combines multiple methods for robust feature selection.
Improved Cuckoo Search Optimization (ICSO)	[87]	No	Yes	Efficient for high-dimensional feature selection problems.
Mann–Whitney U Test	[139]	No	Yes	Non-parametric method for comparing distributions of numerical features across classes.
Entropy-Controlled Firefly Approach (ECfA)	[90]	No	Yes	Guides feature selection using entropy and firefly behavior.

2.3.4 Feature Extraction

Feature extraction is a fundamental preprocessing technique in machine learning, computer vision, and signal processing aimed at transforming raw data into more informative and manageable representations, thereby enhancing model performance and reducing dimensionality [16]. The birth of deep learning has significantly advanced feature extraction, enabling automated, end-to-end learning models that streamline complex analysis and classification tasks. Convolutional neural networks have revolutionized image processing by automatically extracting

complex patterns from images without manual feature engineering, leading to significant application advancements such as object detection and recognition [213]. The effectiveness of feature extraction is ultimately evaluated by its impact on downstream tasks like classification, underscoring its pivotal role in modern data-driven solutions [214]. Table 2.11 summarizes various feature extraction techniques employed in breast cancer studies, detailing their applicability to clinical and/or imaging data and their respective effects.

Table 2.11: Summary of feature extraction techniques and their effects on clinical and imaging data.

Preprocessing Technique	**References**	**Clinical**	**Imaging**	**Effect**
Gray Level Co-Occurrence Matrix (GLCM)	[111]	No	Yes	Extracts texture features (contrast, correlation, energy, homogeneity) by analyzing spatial gray-level relationships, aiding in texture characterization.
Wavelet Multi-Sub-Bands Co-Occurrence Matrix (WMCM)	[215]	No	Yes	Combines wavelet decomposition with GLCM to capture texture features at multiple scales, providing comprehensive texture analysis.
FDCT-WRP	[80]	No	Yes	Integrates fractional discrete cosine transform (FDCT) and wavelet-based region partitioning (WRP) to enhance image detail analysis.
Lifting Wavelet Transform (LWT)	[110]	No	Yes	Decomposes images into frequency sub-bands for multi-resolution analysis, capturing features across different scales.
Hough Transform	[216]	No	Yes	Detects shapes (lines, circles), aiding feature extraction of mass boundaries and calcifications in mammograms.
Discrete Chebyshev Transform (DCST)	[47]	Yes	Yes	Converts images into the frequency domain; low-frequency coefficients are used for feature extraction or dimensionality reduction.
Histogram of Oriented Gradients (HOG)	[100]	No	Yes	Analyzes gradient orientation distributions to capture shape and texture features, effective for lesion characterization in mammograms.
Artificial Intelligent Kit Software	[217]	No	Yes	Utilizes specialized software to extract features from medical images.
Local Binary Pattern (LBP)	[100]	No	Yes	Encodes local pixel relationships into binary patterns, robustly identifying texture patterns like microcalcifications.
Block-Based CDTM and CDF	[87]	No	Yes	Calculates texture features by dividing images into blocks and analyzing intensity contrasts, effectively capturing local texture variations.
Mean Energy	[218]	No	Yes	Measures average intensity or signal strength, characterizing overall image energy or specific frequency bands.
Mean Teager-Kaiser Energy	[218]	No	Yes	Detects rapid changes in signal amplitude and frequency, summarizing edge and transient features in images.
Shannon Wavelet Entropy	[218]	No	Yes	Combines wavelet decomposition with entropy calculation to measure information content or disorder across frequency sub-bands.
Hu's Moments	[114]	No	Yes	Extracts shape descriptors invariant to rotation, scale, and translation, suitable for characterizing breast lesion shapes in ultrasound images.

2.3.5 Images Resizing

Image resizing is a fundamental preprocessing step, particularly crucial for CNNs requiring fixed input dimensions (e.g., 224×224, 256×256) [213]. Standardizing image dimensions enhances training efficiency, optimizes computational resource utilization, and improves memory management, especially when dealing with large datasets [219]. As detailed in Table 2.12, various specific dimensions are employed in breast cancer research [220].

Table 2.12: Summary of image resizing techniques and their effects on imaging datasets.

Preprocessing Technique	References	Clinical	Imaging	Effect
224 × 224	[76, 88, 94, 96, 117, 118, 121, 122, 128, 134, 139, 152, 153, 170, 171, 173, 200, 221–223]	No	Yes	Ensures compatibility, facilitates batch processing, improves computational efficiency, but may lead to loss of fine details, distortion, and interpolation artifacts.
256 × 256	[124, 129, 165, 179]	No	Yes	
299 × 299	[190, 221]	No	Yes	
700 × 460	[160, 167]	No	Yes	
128 × 128	[110, 119]	No	Yes	
64 × 64	[224]	No	Yes	
99 × 99	[225]	No	Yes	
120 × 120	[226]	No	Yes	
250 × 220	[133]	No	Yes	
227 × 227	[227]	No	Yes	
255 × 255	[100]	No	Yes	
296 × 224	[155]	No	Yes	
336 × 224	[85]	No	Yes	
448 × 336	[155]	No	Yes	
448 × 448	[131]	No	Yes	
416 × 416	[184]	No	Yes	
640 × 640	[97]	No	Yes	
1333 × 800	[95]	No	Yes	
Not Specified	[148, 189, 228]	No	Yes	
Padded with Black Background	[217]	No	Yes	Maintains aspect ratio during resizing by padding with a background color.
Multi-scaling	[98, 128, 164, 222]	No	Yes	Captures features at multiple scales using different resolutions.

2.3.6 Cropping

Cropping is a fundamental image processing technique used to remove unwanted areas from an image, focusing attention on the region of interest and improving visual composition and clarity. In the context of medical image analysis, cropping is often used to isolate specific anatomical structures or lesions, reducing the amount of irrelevant data processed by subsequent algorithms. As digital image data continues to grow, efficient cropping methods, especially those incorporating machine learning and computer vision techniques, become increasingly important

for streamlining workflows and enhancing visual quality. Table 2.13 summarizes various cropping techniques used in breast cancer studies and their effects.

Table 2.13: Summary of cropping techniques and their effects on imaging datasets.

Preprocessing Technique	**References**	**Clinical**	**Imaging**	**Effect**
k-means clustering & GMM (Background Elimination)	[229]	No	Yes	Automates background removal using clustering, improving focus on relevant features and model accuracy.
Extraction of region of interest (ROIs)	[80, 87, 94, 110, 111, 134, 166, 195, 200, 215, 225, 226]	No	Yes	Crops images to focus on critical areas (e.g., lesions), reducing irrelevant data and enhancing performance.
Unwanted black/dark parts	[103]	No	Yes	Removes irrelevant background (e.g., dark parts), reducing computational cost and improving accuracy.
Artifacts Removal	[83, 93, 106]	No	Yes	Focuses on breast tissue by removing irrelevant elements, enhancing analysis and diagnosis.
Unimportant regions removal (512×1024)	[105]	No	Yes	Standardizes image size and reduces computational cost for computational efficiency, but risks losing contextual information.
Cropping (128×128)	[230]	No	Yes	
Cropping (512×512)	[82]	No	Yes	
Patches of 224×224	[138]	No	Yes	
	[177]	No	Yes	
	[143]	No	Yes	
Cropping (2048×2048 pixels)	[107]	No	Yes	
Cropping (256×256)	[178]	No	Yes	
Patches of 50×50 pixels	[189]	No	Yes	
Unnecessary and insignificant borders removal	[123]	No	Yes	Removes unnecessary borders to focus on the primary ultrasound image.
Random cropping	[168]	No	Yes	Data augmentation technique to improve model robustness and generalization.
Background (pectoral muscle removal - PMR)	[101]	No	Yes	Improves visibility of breast tissue by removing pectoral muscles in mammograms.

2.3.7 Data Augmentation

Data augmentation is a critical step in machine learning that enhances model performance by artificially expanding the size and diversity of training datasets, mitigating overfitting, and improving the model's ability to generalize effectively. In imaging data, augmentation methods range from basic geometric transformations (e.g., rotation, scaling, flipping, etc.) and photometric adjustments (e.g., brightness changes, color modifications) to more advanced techniques leveraging generative models. The effectiveness of these techniques across various domains highlights the importance of tailoring augmentation strategies to specific tasks to achieve optimal

results [231]. Data augmentation is especially valuable in scenarios with limited training data, as it enhances model robustness and generalization capabilities. The diverse range of augmentation techniques and their effects on clinical and imaging data are presented in Table 2.14.

2.3.8 Filtering

Filtering is vital for improving decision-making and enhancing the quality of training data in machine learning, leading to improved accuracy and reliability. Integrating filtering techniques with deep learning has led to innovative solutions in various applications, highlighting the increasing importance of effective data processing in contemporary data analysis [232]. Various filtering techniques exist, including linear methods like Kalman and moving average filters, which primarily smooth data, and nonlinear methods such as median and Gaussian filters, which are particularly effective in image processing for reducing noise while preserving important edges [233]. Table 2.15 summarizes the filtering techniques employed in breast cancer studies.

2.3.9 Other Preprocessing Techniques

In addition to the core preprocessing methods discussed previously, several other techniques are employed to further refine data for breast cancer diagnosis. These supplementary techniques address specific data characteristics and aim to improve model performance. Table 2.16 summarizes these techniques, detailing their application to clinical and imaging datasets and their specific impacts on data characteristics and subsequent model performance.

Table 2.14: Summary of data augmentation techniques and their effects on clinical and imaging data.

Preprocessing Technique	**References**	**Clinical**	**Imaging**	**Effect**
Rotation	[76,81–83,85,86,88,94, 97,99,100,117,119,122, 126, 128, 129, 131, 134, 138, 148, 151, 152, 154, 155, 158, 159, 161, 165, 166, 170, 171, 174, 189, 190, 201, 203, 218, 225, 227, 234, 235]	No	Yes	Enhances training data diversity by altering image orientation, improving robustness to variations in lesion or anatomical orientations.
Flipping	[76, 81, 83, 85, 86, 88, 94, 96, 97, 100, 117, 119, 122, 126, 128, 129, 131, 133, 134, 151, 152, 154–156, 158, 159, 161, 165, 166, 171, 174, 189, 190, 203, 217, 218, 225, 227, 234, 235]	No	Yes	Mirrors images horizontally or vertically, increasing data diversity and robustness to symmetry-related variations.
Translation (Shifting)	[76,94,97,117,119,122, 138, 148, 159, 170, 171, 174, 189, 203, 225]	No	Yes	Moves images horizontally or vertically, ensuring robustness to variations in object positioning.
Scaling	[82, 86, 117, 128, 148, 151, 158, 174, 189]	No	Yes	Resizes images to handle variations in object size, enhancing model adaptability to magnification differences.
Shearing	[97, 119, 122, 170, 171, 174, 203, 218]	No	Yes	Distorts images by shifting parts relative to others, improving robustness to perspective distortions.
Zooming	[76, 96, 97, 99, 122, 161, 171, 203]	No	Yes	Enlarges portions of images, capturing features at different scales to enhance detection of objects with varying sizes.
Gaussian Blur	[82, 131, 154, 234, 235]	No	Yes	Smooths images to reduce noise, increasing robustness to small variations.
Not specified	[106, 125, 130, 184]	No	Yes	Not specified.
Adding Noise (Gaussian)	[96, 117]	No	Yes	Simulates noise conditions to improve model robustness by diversifying training data.
Adding Noise (Not Specified)	[131, 227]	No	Yes	Increases data variability, improving model robustness to noise. Specific effects depend on noise type.
Cropping	[85, 158, 174]	No	Yes	Extracts smaller regions, focusing on relevant areas and increasing sample variety.
Brightness Changing	[82, 221]	No	Yes	Simulates illumination variations, improving robustness to different lighting conditions.
Multiply by Scalar	[234]	No	Yes	Alters image intensity to simulate variations in exposure or gain.
Color Modification (Reinhard)	[235]	No	Yes	Modifies color distributions to mimic staining variability, improving robustness to imaging variations.
Transpose	[96]	No	Yes	Swaps rows and columns for additional data variety.
Histogram Equalization	[100]	No	Yes	Varies contrast to improve robustness to contrast variations.
Hue/Saturation	[190]	No	Yes	Alters color properties to adapt to staining variations in histopathology.
SMOTE	[168]	No	Yes	Generates synthetic samples for class balancing, improving performance for minority classes.
Color Perturbations	[155]	No	Yes	Introduces variations in brightness, contrast, hue, and saturation for robustness to color inconsistencies.
Contrast Limited Adaptive Histogram Equalization (CLAHE)	[98]	No	Yes	Enhances local contrast, creating alternative image versions.
Median Filter	[98]	No	Yes	Reduces noise and creates slight image variations, similar to Gaussian Blur.

Table 2.15: Summary of filtering techniques and their effects on imaging and clinical data.

Preprocessing Technique	**References**	**Clinical**	**Imaging**	**Effect**
CLAHE	[47, 88, 90, 98, 101, 106, 188, 195, 226, 236]	No	Yes	Enhances contrast, improves detail visibility, limits noise amplification.
Median filter	[87, 98, 152, 226, 228]	No	Yes	Reduces impulse noise, preserves edges.
Adaptive histogram equalization (AHE)	[81, 93, 103, 194]	No	Yes	Enhances local contrast, amplifies noise (often replaced by CLAHE).
Gaussian-blurring filter	[152, 179, 188]	No	Yes	Smooths images, reduces noise, retains structural details.
Adaptive median filter	[103, 229]	No	Yes	Removes impulse noise, preserves fine details via adjustable neighborhood.
Mean filter	[161, 228]	No	Yes	Smooths images, may blur edges.
Fuzzy non-local means (NLM)	[91, 237]	No	Yes	Reduces noise using fuzzy logic, preserves fine details.
Laplacian of Gaussian (LoG) filter	[130, 201]	No	Yes	Detects edges, enhances intensity changes, highlights lesion boundaries.
Wiener filter	[101, 224]	No	Yes	Minimizes noise using power spectrum, restores image quality.
Histogram Equalization	[104, 111, 124, 167, 188, 202]	No	Yes	Improves contrast, may amplify noise.
Difference filter	[82]	No	Yes	Highlights intensity changes for edge detection.
Hough transform	[142]	No	Yes	Detects lines/shapes for boundary identification.
Sharpening	[157]	No	Yes	Enhances edges/details, may amplify noise.
Gamma correction	[157]	No	Yes	Adjusts brightness/contrast for better tissue visualization.
Multi-threshold Peripheral Equalization	[81]	No	Yes	Enhances contrast in peripheral mammogram regions.
Adaptive Fuzzy median Filter	[202]	No	Yes	Reduces noise, preserves details using fuzzy logic and adaptive techniques.
Holistically Nested Edge Detection (HED)	[236]	No	Yes	Produces precise edge maps for cell boundary delineation.
Weighted filtering (WF)-based Noise Removal	[127]	No	Yes	Reduces noise with selective smoothing, preserves edges.
Fuzzy Histogram Equalization (FHE)	[188]	No	Yes	Enhances contrast with reduced artifacts using fuzzy logic.
Anisotropic Diffusion Filtering	[225]	No	Yes	Smooths homogeneous areas, preserves edges.
Reducing Anisotropic Diffusion (SRAD)	[225]	No	Yes	Faster anisotropic diffusion, suitable for ultrasound images.
Adaptive Unsharp Mask Filters (AUMF)	[223]	No	Yes	Enhances edges with controlled sharpening.
Multiscale principal component analysis (MPCA)	[218]	No	Yes	Reduces noise by decomposing images and applying PCA.
Wavelet denoising	[202]	No	Yes	Removes noise across frequency bands, retains details.
RandomRBF	[48]	Yes	No	Transforms features to higher dimensions for better ML performance.

Table 2.16: Summary of other preprocessing techniques and their effects on clinical and imaging data.

Preprocessing Technique	**References**	**Clinical**	**Imaging**	**Effect**
Feature Weighting (K-means clustering features)	[46]	Yes	No	Improves classification accuracy by emphasizing key clusters and reducing noise.
Kernel Neutrosophic C-Means Clustering (KNCM)	[30]	Yes	No	Enhances robustness and accuracy for uncertain or overlapping data.
Image Binarization (e.g., OTSU, Niblack)	[109]	No	Yes	Highlights regions of interest and aids segmentation.
Not Specified	[93]	No	Yes	Simplifies images, potentially highlighting key features.
Data Generation (VAE)	[28]	Yes	No	Addresses class imbalance and enhances model robustness.
Data Generation (GAN)	[28,34]	Yes	No	Addresses class imbalance and improves training robustness.
Morphological Operations (e.g., erode, dilate)	[105,216,228]	No	Yes	Removes noise, smooths edges, and enhances lesion boundaries.
Color Space Conversion (RGB to HSV, LAB, etc.)	[161,181]	No	Yes	Improves robustness to staining or brightness variations.
Three-Binary Classes Conversion	[32]	Yes	No	Allows granular analysis but increases complexity.
RGB to Grayscale Conversion	[79, 84, 134, 230]	No	Yes	Reduces computational complexity by removing color information.
Grayscale to RGB Conversion	[124,139]	No	Yes	Ensures compatibility with RGB-based architectures.
Color Channel Extraction	[181]	No	Yes	Reveals subtle features specific to certain color components.
Three-Channel Image Construction	[138]	No	Yes	Enhances feature learning from complementary data.
Image Conversion (1D to 2D)	[74]	Yes	No	Enables spatial analysis using image-based techniques.

2.4 Machine Learning-Based Classification Methods

Machine learning is an essential branch of AI that enables computers to identify patterns in data and make predictions without explicit programming. Among its four primary types, Supervised Learning (SL), Unsupervised Learning (USL), Semi-Supervised Learning (SSL), and Reinforcement Learning (RL), SL is particularly relevant for tasks that require labeled data, such as predictive analytics and healthcare applications, by mapping input features to known outputs [238, 239]. However, when encountering inputs deviating from training data, SL struggles with generalization. Given their dominance in the literature, this section concentrates on the most frequently used supervised ML techniques for breast cancer detection and diagnosis. The availability of labeled healthcare datasets, where target variables (e.g., diagnosis or tu-

mor type) are predefined, makes supervised learning especially suitable for this domain. These techniques consistently achieve high accuracy in classification tasks, playing a critical role in advancing data-driven healthcare solutions. Figure 2.2 illustrates the trends in research papers employing various SL techniques for breast cancer classification from 2018 to 2024, reflecting the research community's evolving preferences and innovations.

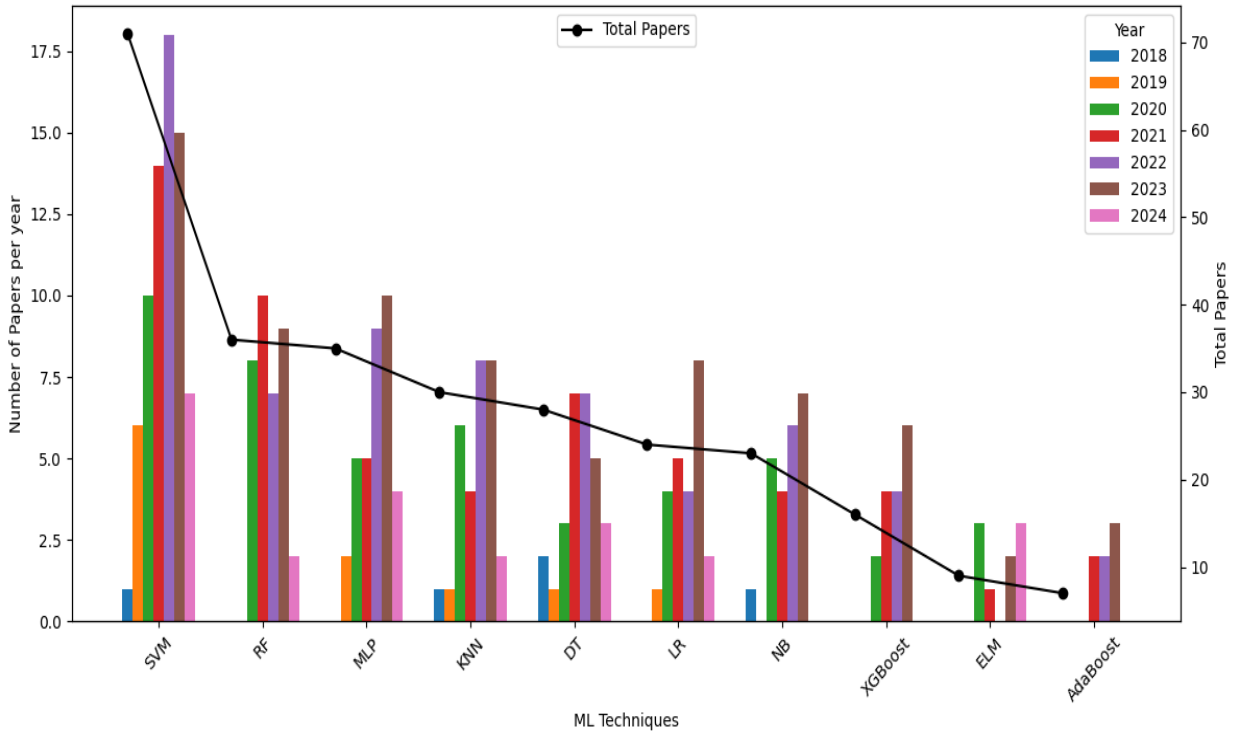


Figure 2.2: Most frequently used supervised machine learning techniques for breast cancer classification (2018–2024).

2.4.1 ML-Based Techniques for BC Diagnosis With Clinical Data

A comprehensive analysis of ML techniques applied to clinical datasets for breast cancer diagnosis is presented in Table 2.17, offering a detailed comparison of various methods, including Artificial Neural Networks (ANNs), Radial Basis Functions (RBFs), and sophisticated hybrid models such as CWV-BANN-SVM. These models were assessed using well-established datasets, including WBCD, WDBC, WPBC, SEER, and Coimbra. Several approaches have demonstrated exceptional performance, like CWV-BANN-SVM, reporting perfect (100%) accuracy in specific studies. This comparative analysis not only presents key performance metrics (accuracy, sensitivity, specificity, precision, F1-score, and Area Under the Curve (AUC), where available) for

each model but also summarizes the core novelty of the corresponding research, offering valuable insights into their potential to enhance diagnostic accuracy and ultimately improve patient outcomes.

Chaurasia et al. [56] compared NB, RBF Network, and DT for breast cancer diagnosis using the WBCD dataset. Their analysis, employing 10-fold cross-validation, found that NB achieved the highest accuracy (97.36%), sensitivity (97.4%), and specificity (97.9%), outperforming the RBF Network and DT. The study emphasized NB's predictive power and suggested further validation with diverse datasets.

Abdar et al. [49] introduced CWV-BANN-SVM, a novel technique for breast cancer prediction that combines SVMs and Boosted ANNs (BANN) using Confidence-Weighted Voting (CWV). Applying the WBCD, the model optimized SVM parameters and achieved perfect performance with an accuracy of 100% using a 50/50 training/testing split. This highly accurate model shows promise for effective computer-aided breast cancer diagnosis and other disease prediction tasks.

Memon et al. [39] presented an ML-based system for breast cancer detection using SVM and Recursive Feature Elimination (RFE) algorithms on the WDBC dataset. Preprocessing and feature selection were done before training and testing the model. The SVM with a linear kernel achieved 99% accuracy, specificity, and 98% sensitivity on 18 key features. The system outperformed previous methods, highlighting RFE's effectiveness. This system could enhance early and accurate breast cancer diagnosis in healthcare settings.

Hajiabadi et al. [46] developed a novel ML technique for breast cancer prediction that uses an ensemble loss function within an artificial neural network. The method improves robustness and probability estimation by integrating three loss functions—correntropy, hinge, and cross-entropy. It achieves joint optimization of loss weights and network parameters via backpropagation. Evaluated on the WBCD dataset, this method reached an accuracy of 97%, outperforming individual loss function models and other classifiers like KNN and SVM, even with label noise up to 30%.

Osman et al. [55] developed an automated breast cancer classification system using an Ensemble Boosting Learning (EBL) method combined with a RBF neural network (RBF-EBL). They utilized four UCI datasets and features like cell size and texture, applying 10-fold cross-validation. This approach improved diagnostic accuracy to 98.4% on the WBCD dataset, outperforming other classifiers and showing promise for early breast cancer detection and treatment.

Ak et al. [43] analyzed various supervised learning algorithms for breast cancer classification using the WDBC dataset. They found that LR achieved the highest accuracy at 98.06%,

with KNN and SVM also performing well. RF and DT showed variability due to random feature selection, while NB performed poorly. The study emphasized the importance of FS and algorithm choice for optimal classification, especially with unbalanced datasets.

Al-Azzam and Shatnawi [44] evaluated supervised learning versus semi-supervised learning algorithms for breast cancer diagnosis using the WDBC dataset. They applied nine classification algorithms, including LR and KNN, revealing similar performance between SL and SSL. LR achieved 97% accuracy, while KNN reached 98%. Their findings indicate that SSL can effectively use unlabeled data, potentially reducing the need for costly labeled datasets while preserving high accuracy in breast cancer diagnosis.

Gopal et al. [48] proposed an Internet of Things (IoT)-enabled breast cancer diagnosis system that utilizes machine learning techniques on the WBCD. After data normalization and filtering, PCA reduced the feature set to 11 components. Classification was done using RF and Multi-Layer Perceptron (MLP) classifiers with 10-fold cross-validation. Results showed that MLP outperformed RF, achieving 98% accuracy compared to RF's 95%, with better precision, recall, and F1-score. The system incorporates IoT devices for tumor detection and data acquisition. This integrated approach aims to improve early breast cancer diagnosis and potentially lower mortality rates.

Sharma et al. [24] introduced a hybrid feature ensemble learning method, named Neural Network-Extra Tree (NN-ET), for breast cancer classification using the WDBC dataset. The model combines a feature reduction neural network with an extra tree classifier for decision-making. It achieved the highest accuracy of 99.74% and outperformed various classifiers, including LR, SVM, KNN, RF, and ensemble methods like Adaboost and Gradient Boosting (GB). The study concludes that the NN-ET model is highly effective in distinguishing between benign and malignant breast cancer.

Arowolo et al. [35] used machine learning for breast cancer diagnosis using the WDBC dataset. They applied the LDA for feature extraction and tested RF and SVM classifiers with an 80/20 training/testing split. The SVM achieved a higher accuracy of 96.4% compared to RF's 95.6%. The study highlights the potential of ML in breast cancer diagnosis and suggests further research into feature extraction techniques to enhance predictive accuracy.

Nanglia et al. [60] introduced a breast cancer prediction method using a heterogeneous ensemble model that incorporates KNN, SVM, and DT as base classifiers, along with a RF model and four single classifiers, including Stochastic Gradient Descent (SGD), LR, ANN, NB. Using the Coimbra Breast Cancer dataset (116 instances, 10 predictors) and following the CRISP-DM methodology, the study used data visualization techniques and evaluated performance with

k-fold cross-validation (70/30 split). The obtained results showed that the ensemble model significantly outperformed individual classifiers, achieving the highest accuracy and demonstrating the effectiveness of ensemble methods in breast cancer prediction.

Rashed et al. [31] evaluated seven automated machine learning (AutoML) techniques: Orange, Lazy Predict, TPOT, MLJAR, MATLAB Classification Learner, and AutoKeras—for breast cancer diagnosis using eight datasets. They found that Lazy Predict and MATLAB Classification Learner excelled in binary classification tasks, offering guidance for selecting effective models.

Strelcenia and Prakoonwit [34] addressed the issue of limited breast cancer detection datasets by introducing a Kullback-Leibler Divergence Conditional Generative Adversarial Network (K-CGAN) to generate synthetic data. Utilizing the Wisconsin Breast Cancer dataset (357 malignant, 212 benign cases), they trained several classification models, including XGBoost, RF, KNN, MLP, and LR. The MLP model outperformed others with 98.52% precision, 100% recall, 99.26% F1-score, and 99.25% accuracy, highlighting the effectiveness of GAN-based data augmentation in improving classification performance in imbalanced medical datasets.

Wu et al. [63] conducted a retrospective study using machine learning to predict overall survival in Second Primary Breast Cancer (SPBC) patients, utilizing data from the SEER Program. They analyzed 10,321 SPBC patients and validated their models on an independent cohort of 3,638 patients. Four machine learning models—Random Survival Forest (RSF), GB, Extremely Randomized Trees (EST), and Survival Tree (ST)—were compared to the traditional Cox proportional hazards model. The RSF model outperformed the others, achieving a time-dependent area under the ROC curve (t-AUC) of 0.805 and an integrated Brier score (iBrier) of 0.123 in the test set, indicating its potential to improve SPBC clinical management.

2.4.2 ML-Based Techniques for BC Diagnosis With Imaging Data

A comparative overview of ML techniques applied to imaging datasets for breast cancer diagnosis is presented in Table 2.18. This overview encompasses a range of methods, including SVMs, Extreme Learning Machines (ELMs), XGBoost, and various ensemble models, evaluated using datasets such as MIAS, DDSM, BUSI, and BreakHis. Notably, models like DE-RBF-KELM and MFO-ELM (combined with LWT, PCA, and LDA for feature processing) achieved perfect accuracy on the MIAS dataset while maintaining respectable performance on other datasets. This variability in model performance across different datasets highlights the importance of careful data selection and model optimization.

Table 2.17: Performance comparison of machine learning techniques for breast cancer diagnosis using clinical datasets.

Refs	**Year**	**Dataset(s)**	**Method(s)**	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**Precision (%)**	**F1-score (%)**	**AUC (%)**
[56]	2018	WBCD	NB	97.36	97.4	97.90	-	-	-
[49]	2019	WBCD	CWV-BANN-SVM	100.0	-	-	-	-	-
[39]	2019	WDBC	Linear SVM	99.0	98.0	99.0	-	99.0	-
[55]	2020	WDBC, WBCD	RBFNN	98.40	-	-	-	-	-
[43]	2020	WDBC	LR	98.06	-	-	-	-	-
[46]	2020	WBCD	ANN	97.0	94.0	-	100.0	97.0	-
[48]	2021	WBCD	MLP	98.0	97.0	-	98.0	96.0	-
[44]	2021	WDBC	KNN	98.0	98.0	-	98.0	98.0	-
[35]	2022	WDBC	LDA + SVM	96.4	95.7	97.8	96.4	97.8	-
[60]	2022	Coimbra	Heterogeneous Ensemble Model (KNN, DT, SVM)	78.0	-	-	-	-	-
[24]	2022	WDBC	NN-ET	99.74	100.0	99.06	98.46	99.21	-
[31]	2023	WDBC, WBCD, WPBC, SEER, Coimbra	MLJAR, Lazy Predict	99.12	-	-	-	-	-
[34]	2023	WDBC	XGBoost with K-CGAN	99.12	100.0	-	98.28	99.13	-
[63]	2024	SEER	Random Survival Forest (RSF)	-	-	-	-	-	80.5

Vijayarajeswari et al. [216] developed a system for the early detection and classification of breast cancer mammograms. Their approach utilized the Hough transform for feature extraction, effectively identifying specific shapes within preprocessed mammogram images, which involved noise reduction and label removal. Employing an SVM for classification, they analyzed a dataset of 95 mammograms from the MIAS database, categorized as normal, benign, and malignant, achieving an impressive accuracy of 94%, underscoring its effectiveness in distinguishing between normal and abnormal mammograms.

Muduli et al. [110] designed a CAD method for classifying breast masses effectively. Their method employs LWT for feature extraction, followed by PCA and LDA for dimensionality reduction, and an optimized ELM using Moth-Flame Optimization (MFO) for Classification. Using 5-fold cross-validation, the model achieved perfect accuracy of 100% on the MIAS dataset, demonstrating greater computational efficiency and superior accuracy with fewer features than

existing methods.

AlFayez et al. [194] devised a breast cancer detection system using thermograms, comparing ELM and MLP classifiers. Their approach included preprocessing for noise reduction, k-means clustering for segmenting the region of interest and extracting geometric and textural features. Utilizing the DMR-IR dataset with 1,345 thermograms, they found that ELM with the TRIBAS activation function outperformed MLP, achieving the highest classification accuracy of 99.10% compared to MLP's 82.20%. It demonstrated ELM's effectiveness for thermogram-based breast cancer detection, enabling rapid and accurate diagnosis.

Alì et al. [240] created a Decision Support System (DSS) that integrates BI-RADS classification with radiomic classifiers to lower false positives (FPs) in breast calcifications detected through digital breast tomosynthesis (DBT). Analyzing a dataset of 49 patients, they extracted pertinent features and trained a machine learning classifier using nested 10-fold cross-validation. This approach led to a 50% reduction in FPs, doubled the positive predictive value (PPV), and achieved 82% accuracy, 78% sensitivity, 85% specificity, and an AUC of 80%.

Muduli et al. [80] introduced a computer-aided diagnosis system to classify breast masses as benign or malignant. Their approach employs the Fast Discrete Curvelet Transform with Wrapping (FDCT-WRP) for feature extraction, followed by PCA and LDA for dimensionality reduction. They used an ELM optimized with Particle Swarm Optimization (MODPSO-ELM) for the classification stage, achieving accuracy rates of 100%, 98.94%, and 98.76% on the MIAS, DDSM, and INBreast datasets, respectively, surpassing the performance of conventional ML methods such as SVM and KNN.

Darweesh et al. [111] proposed a hierarchical ML system for early breast cancer diagnosis based on the MIAS mammography dataset. First, the two-level classification differentiates normal from abnormal cases with Gray-Level Co-occurrence Matrix (GLCM) features and a RF classifier, achieving 97% accuracy. Abnormal cases are then classified as benign or malignant using Local Binary Patterns (LBP), yielding 75% accuracy. The overall system achieved 85% classification accuracy, highlighting this staged approach's effectiveness.

Jebarani et al. [229] suggested a hybrid method for breast cancer detection in digital mammography, combining k-means clustering with a Gaussian Mixture Model (GMM). Their approach included preprocessing steps like adaptive median filtering for noise reduction and achieved the highest accuracy of 95.5%. This method improved the classification of benign and malignant tumors and supported early breast cancer diagnosis, aiding radiologists in better identifying malignant regions.

Khairnar et al. [109] investigated the impact of different image binarization thresholds

on breast cancer identification in mammography images. Evaluating methods such as Otsu, Niblack, Bernsen, and Thepade's Sorted Block Truncation Coding (TSBTC) for converting grayscale images to binary representations for region isolation, they extracted features to train various machine learning classifiers. Notably, the REPTree classifier, when used with Otsu thresholding, achieved the best individual performance at 99.68%. The study concluded that Otsu thresholding is the most effective method for isolating malignant regions, significantly influencing the accuracy of breast cancer identification.

Kaur et al. [65] presented Bayesian hyperparameter optimized Stacked Ensemble (BSense), a model predicting breast cancer survival on the genomic dataset. BSense combines DNNs, Gradient Boosting Machines(GBMs), and Distributed Random Forests (DRFs), optimizing hyperparameters through parallel Bayesian Optimization (BO) and utilizing the Artificial Bee Colony (ABC) algorithm for feature selection. Tested on diverse datasets from The Cancer Genome Atlas (TCGA), METABRIC, metabolomics, and RNA-seq, BSense achieved an AUC of 83.9% and 85.2% accuracy on the TCGA dataset. This ensemble model shows promise for enhancing patient-specific treatment plans.

Bacha et al. [47] developed a breast cancer diagnosis system using an RBF-KELM optimized by Differential Evolution (DE). By optimizing key parameters, they improved classification performance, achieving 100% accuracy on the MIAS dataset. The DE-RBF-KELM outperformed SVM, Poly-KELM, and other methods, particularly excelling in sensitivity, specificity, and the AUC, with a combined loss function that enhanced robustness against outliers.

Nastase et al. [114] analyzed machine learning for distinguishing malignant and benign breast masses in the BUSI dataset. They extracted shape features using Hu's moments and trained a KNN classifier and an RBFNN. The KNN classifier achieved 85% and 80% accuracy with Hu's moments M1 and M5, while the RBFNN reached 76% accuracy with moment M1. This study highlights the effectiveness of image segmentation and shape feature extraction in breast cancer diagnosis using the BUSI dataset.

Hussain et al. [79] developed a hybrid feature extraction approach integrating various feature methods, such as texture, morphological features, SIFT, GLCM features, entropy, elliptic fourier descriptors, and Reconstruction-Independent Component Analysis (RICA) for automated breast cancer detection, focusing on the complex characteristics of masses and microcalcifications. Combining RICA with textural features using an SVM with a gaussian kernel obtained the highest performance, achieving an accuracy of 97.55% and a ROC-AUC of 0.9976, compared to other machine learning classifiers like DT, KNN, and NB, highlighting its potential for early breast cancer detection.

Liu et al. [215] built an ML-based framework to predict breast cancer recurrence and metastasis risk using features from histopathological images. They used XGBoost, RF, and LR on data from 127 patients at the Cancer Hospital, Chinese Academy of Medical Sciences. They were validated with 88 samples from the TCGA dataset. The XGBoost model achieved AUC values of 0.75, identifying key predictive features involving the second moment of the B color component and the detail level mean square error of the wavelet multi-sub-bands co-occurrence matrix. This tool provides reliable risk assessments, potentially reducing pathologists' workload and improving patient survival outcomes.

Liang et al. [212] employed the XGBoost algorithm to classify breast cancer based on mammograms exhibiting microcalcifications to assist oncologists and medical image processing engineers in differentiating benign and malignant lesions based on a dataset of calcification images from a hospital. Focusing on 51 initial features, they selected the top 15 most relevant features for final analysis. XGBoost achieved the highest accuracy of 90.24%, outperforming the other machine learning techniques, including KNN, DT, AdaBoostM1, RF, and Gradient Boosting Decision Tree (GBDT), demonstrating a lower prediction error rate, as validated by the Kolmogorov-Smirnov predictive accuracy test.

Haider et al. [146] introduced the Penalized Sequential Discriminative Dictionary Learning (PSDDL) framework for breast cancer classification, targeting imbalanced datasets. By combining class-specific dictionary learning with sparse coding, PSDDL constructs structured dictionaries. Using the BreakHis dataset, which features histopathological images at various magnifications, the method achieved a 97.52% mean classification accuracy at 40x magnification, surpassing existing techniques and effectively addressing class imbalance while enhancing feature representation. The study emphasizes PSDDL's potential for automated cancer diagnosis in medical image analysis.

2.5 Deep Learning-Based Classification Methods

Deep learning is a subfield of ML that uses multi-layered artificial neural networks to extract complex patterns from various data types, such as images and numerical data [232]. This ability to process unstructured data has led to significant advancements in bioinformatics and computer vision applications, often outperforming traditional ML algorithms [241]. Convolutional neural networks [219] have become essential in image-based tasks, particularly in medical imaging for breast cancer diagnosis. They consist of several layers: convolutional layers extract features with filters, pooling layers reduce dimensionality while preserving information, activation functions

(like sigmoid and ReLU) determine neuron activation, and fully connected layers link neurons for final classification. CNNs excel with large datasets and perform strongly in various diagnostic applications [232].

Table 2.18: Performance comparison of machine learning techniques for breast cancer diagnosis using imaging datasets.

Refs	**Year**	**Dataset(s)**	**Method(s)**	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**Precision (%)**	**F1-score (%)**	**AUC (%)**
[216]	2019	Mammograms	SVM	94.0	-	-	-	-	-
[194]	2020	DMR-IR	ELM	99.10	97.03	98.05	-	-	-
[110]	2020	MIAS, DDSM	LWT + PCA + LDA + MFO-ELM	100.0	-	-	-	-	100.0
[240]	2021	DBT images	Ensemble of 200 Decision Trees with Gini index	82.0	78.0	85.0	-	-	80.0
[80]	2021	MIAS, DDSM	MFO-ELM	99.76	-	-	-	-	-
[111]	2021	MIAS	Hierarchical-based classifier	97.0	-	-	-	-	-
[229]	2021	MIAS	K-means + GMM	95.50	-	-	-	-	-
[109]	2021	MIAS	REPTree with Otsu thresholding	99.68	-	-	-	-	-
[114]	2022	BUSI	Huś moment + K-NN	89.0	83.0	-	87.0	83.0	-
[79]	2022	DDMS	SVM Gaussian	97.55	-	-	-	-	99.76
[215]	2022	TCGA FFPE	XGBoost	94.0	-	-	-	-	-
[47]	2022	MAIS	DE-RBF-KELM	100.0	100.0	100.0		100.0	-
[65]	2022	TCGA-BRCA	BSense	84.2	-	-	-	-	-
[212]	2022	Mammograms	XGBoost	90.24	88.45	-	90.00	89.52	89.03
[146]	2023	BreakHis	PSDDL	97.4	97.2	-	97.6	97.4	-

Transfer learning has revolutionized DL by enabling the reuse of pre-trained CNN models like AlexNet, VGGNet, and ResNet, which are primarily trained on large datasets such as ImageNet [242], reducing training time and complexity while enhancing classification accuracy, which is particularly beneficial in medical imaging where labeled data is limited [243]. Meanwhile, Vision Transformers (ViTs) offer a promising alternative to CNNs by utilizing attention mechanisms to capture global context and spatial relationships in complex medical images [244]. Furthermore, fine-tuning and fusion methods are increasingly used to enhance the robustness and effectiveness of hybrid systems [245].

This section offers concise descriptions of commonly used DL models in the classification and diagnosis of breast cancer, including ResNet, VGGNet, and AlexNet, considered foundational tools in numerous studies. Figure 2.4 illustrates the range of DL techniques discussed in the literature, and Figure 2.3 presents the trends in research publications utilizing various DL methods for breast cancer classification from 2018 to 2024, highlighting the evolving preferences and ongoing innovations within the research community.

2.5.1 DL-Based Techniques for BC Diagnosis With Clinical Data

This subsection analyzes the application of DL-based techniques to clinical datasets for breast cancer diagnosis, as summarized in Table 2.19, including customized CNNs, pre-trained models, and hybrid approaches such as CNNs combined with PSO (CNN+PSO) and LSTM networks combined with Gated Recurrent Units (LSTM+GRU). This analysis underscores the effectiveness of DL models in clinical breast cancer diagnosis while highlighting areas for further research and development. However, challenges remain, including the requirement for large, high-quality training datasets, the significant computational resources needed for complex models, and the limited generalization capabilities often observed across different datasets.

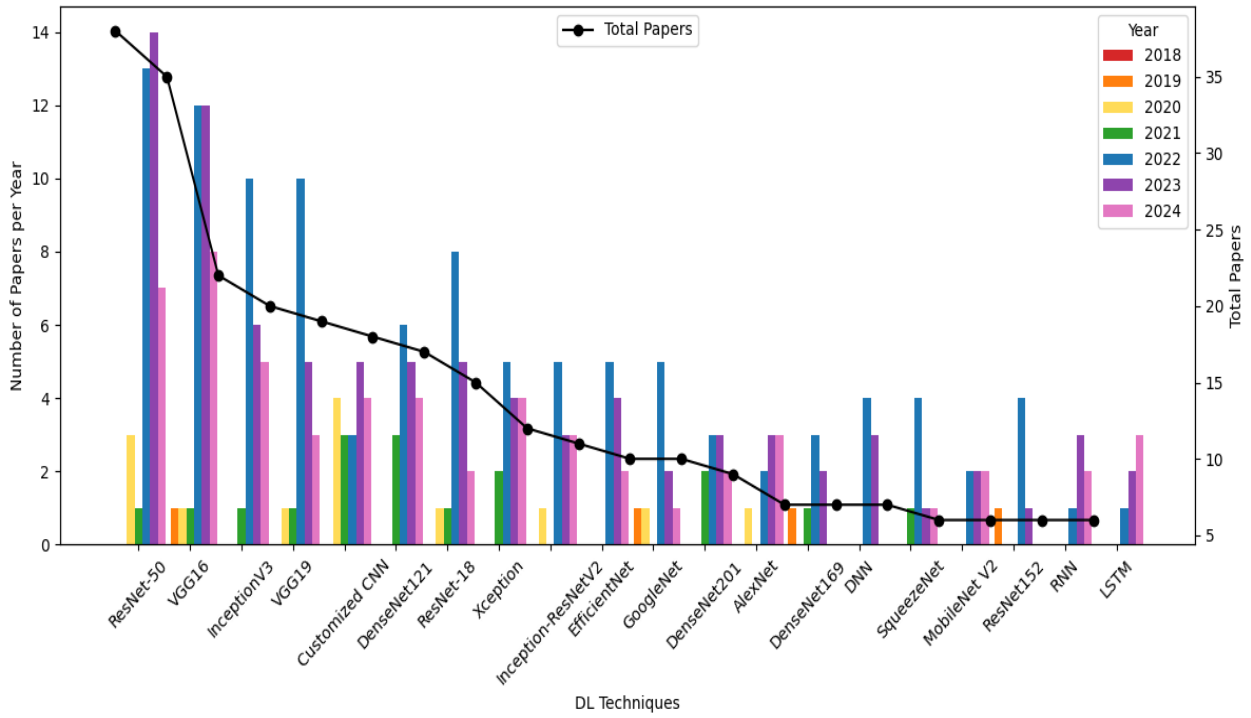


Figure 2.3: Most frequently used deep learning techniques for breast cancer classification (2018–2024).

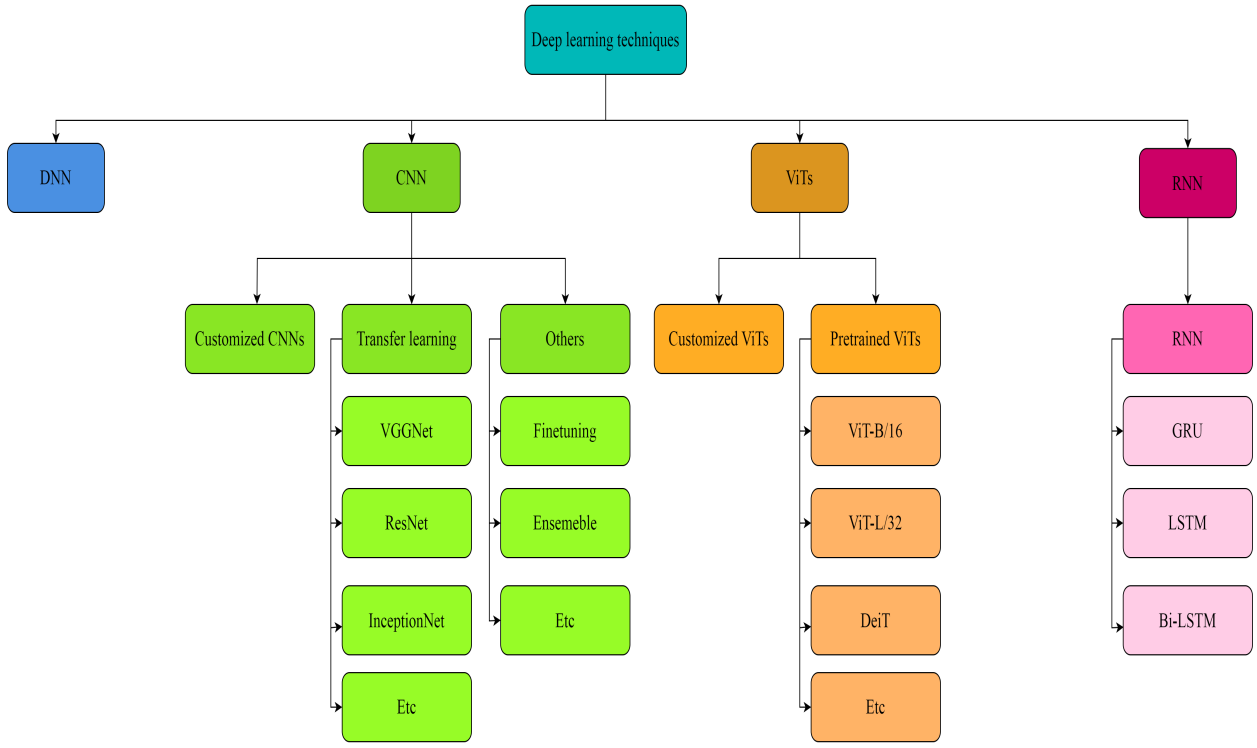


Figure 2.4: A comprehensive overview of deep learning techniques.

Arya and Saha [68] developed a DL-stacked ensemble model for breast cancer prognosis prediction using multi-modal data from 1,980 patients in the METABRIC dataset, including gene expression and copy number alteration profiles. Their two-stage approach involves using CNNs for feature extraction, followed by a stacked ensemble model with RF and SVM classifiers. This model achieved an AUC of 0.93 and 90.2% accuracy, surpassing existing methods.

Arya et al. [67] introduced the Stacked Generalized Attention CNN (SiGaAtCNN) for feature extraction from multi-modal data in breast cancer survival prediction. Combining SiGaAtCNN with Stacked Random Forest (SRF), the proposed approach achieved an AUC of 0.950 and 91.2% accuracy on the METABRIC dataset, outperforming other comparative models. This study underscores the potential of gated attention CNNs and ensemble methods to enhance breast cancer survival prediction, with promising implications for personalized treatment strategies.

Das et al. [74] proposed an ensemble DL method for breast cancer detection that converts one-dimensional gene expression data into images using t-SNE and Convex Hull algorithms. Their two-stage classification employs three CNNs as base classifiers, followed by an MLP. The method also includes Empirical Wavelet Transform (EWT) and Variational Mode Decomposition (VMD) for dataset decomposition, showing superior performance compared to existing methods.

Ogundokun et al. [199] proposed a framework to optimize hyperparameters in CNNs and ANNs for improving breast cancer detection. Using the WDBC dataset, they applied PSO for tuning and compared SVM and MLP classifiers. The ANN model achieved 99.2% accuracy, slightly higher than CNN's 98.5%, indicating that ANNs can perform exceptionally well on smaller datasets. The study recommends further research on advanced neural architectures and data security in IoT systems.

Othman et al. [70] created a hybrid deep learning framework using decision-level fusion to predict breast cancer survival with multi-omics data from the METABRIC dataset. The model employs a CNN for feature extraction, followed by LSTM and GRU classifiers, achieving accuracies of 97.0% and 97.5%, respectively. By the fusion of the decision, the combined approach reached 98.0% accuracy, highlighting the benefits of integrating multi-modal data for enhanced prediction accuracy.

Chakravarthy et al. [103] proposed a novel breast cancer classification framework using deep learning and metaheuristic optimization. Their approach utilizes ResNet18 for feature extraction and the Crow Search Optimization Algorithm (CSOA) for feature transformation, followed by classification with a Weighted KNN (wKNN) classifier optimized through Grid Search (GS). The CSOA-wKNN model achieved accuracies of 84.35% on MIAS, 83.19% on INBreast, and 97.36% on WDBC, surpassing other methods and demonstrating advanced performance in breast cancer classification.

Asadi et al. [104] introduced an efficient cascaded network for breast cancer diagnosis, using UNet for tumor segmentation and a streamlined ResNet50-based model for classification. The segmentation achieved an F1-score of 97.30%, while the classification reached an accuracy of 98.61% and an F1-score of 98.41%, highlighting the proposed method's effectiveness in diagnosing breast cancer.

2.5.2 DL-Based Techniques for BC Diagnosis With Imaging Data

A comparison of DL-based techniques for breast cancer diagnosis is summarized in Tables 2.20 and 2.21. It reviews models like ResNet, DenseNet, and CNN-BiLSTM using datasets such as BUSI and BreaKHis, demonstrating DL models' potential in medical image analysis. However, challenges remain, including the need for dataset-specific tuning and difficulties in generalizing models across different imaging modalities. The table also highlights the key contributions and novelties of each study.

Table 2.19: Performance comparison of deep learning techniques for breast cancer diagnosis using clinical datasets.

Refs	Year	Dataset(s)	Method(s)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)	AUC (%)
[68]	2020	METABRIC	CNNs with a stacked ensemble model (RF and SVM)	88.1	95.0	95.0	94.9	-	96.8
[67]	2021	METABRIC	SiGaAtCNN + RF	93.3	87.2	-	84.5	-	92.9
[74]	2021	1D gene exp.	Stacked DL	98.08	99.20	93.55	98.41	98.80	-
[199]	2022	WDBC	CNN+PSO+MLP	97.2	97.8	96.3	97.8	97.6	-
[70]	2023	METABRIC	Hard voting (LSTM + GRU)	98.0	99.2	-	99.0	-	98.2
[103]	2023	WDBC	ResNet18 with CSOA transformed wKNN	79.13	74.51	82.81	77.55	76.00	-
[104]	2023	WDBC	ResNet50	96.8	96.8	-	96.8	96.8	-

Yang et al. [155] introduced Ensemble of MultiScale convolutional neural Networks (EMS-Net) to effectively classify breast histopathological images into four categories: normal, benign, in situ carcinoma, and invasive carcinoma. Their model used a multi-scale strategy with data augmentation and fine-tuning of pre-trained models (DenseNet-161, ResNet-152, and ResNet-101) on image patches. By employing an ensemble approach and majority voting, they achieved a mean accuracy of 91.75% $\pm$ 2.32% in five-fold cross-validation on the BACH dataset and 90.00% in online verification, outperforming existing methods.

Chougrad et al. [94] proposed a novel approach for early breast cancer diagnosis using multi-label transfer learning with CNNs, evaluated on four benchmark datasets (CBIS-DDSM, BCDR, INBreast, and MIAS). Their method employs joint learning through multi-label image classification and introduces a novel fine-tuning strategy to adapt pre-trained CNNs to the specific task. They created a customized label decision scheme to enhance the confidence levels for each visual concept, demonstrating superior performance compared to commonly used baseline methods.

Al-antari et al. [81]developed a CAD system for breast lesions using deep learning on the DDSM and INbreast datasets. They used a YOLO detector for lesion detection and modified three deep learning models: a standard CNN, ResNet-50, and InceptionResNet-V2, integrating global average pooling and softmax layers. The system achieved detection accuracies of 99.17% on DDSM and 97.27% on INbreast, with classification accuracies of 94.50% (CNN), 95.83% (ResNet-50), and 97.50% (InceptionResNet-V2) on DDSM, and 88.74%, 92.55%, and 95.32% on INbreast, showing strong diagnostic performance.

Wang et al. [147] introduced Prototype Transfer Generative Adversarial (PTGAN), an unsupervised learning method that combines GANs and prototypical networks for domain adaptation. It utilizes a small labeled source dataset to train a model for a larger, unlabeled target dataset, reducing the distribution gap between images from different devices. By embedding learned feature vectors into a metric space, PTGAN enhances target domain classification. Experiments on the BreaKHis dataset showed PTGAN achieved nearly 90% accuracy in classifying benign and malignant tissues, demonstrating its effectiveness and cost-efficiency for clinical breast cancer classification.

Boumaraf et al. [153] introduced a transfer learning approach for breast cancer classification using histopathological images from the BreaKHis dataset. They tackled both magnification-dependent and independent tasks, utilizing a fine-tuned ResNet-18 architecture. They applied Global Contrast Normalization (GCN) and a three-fold data augmentation strategy to improve performance and generalization, showing enhanced effectiveness across classification scenarios.

Aljuaid et al. [152] developed a CAD system for breast cancer classification using deep neural networks and transfer learning. They utilized ResNet18, ShuffleNet, and Inception-V3Net architectures based on the BreakHis dataset. The results showed that ResNet18 surpassed other models, achieving the highest accuracy of 99.7%. These findings demonstrated the model's capacity to improve the performance of breast cancer detection and classification.

Karthiga et al. [106] developed an AI-based breast cancer diagnosis system using transfer learning with CNNs and hyperparameters tuning for mammogram analysis. The system was tested on the MIAS, DDSM, and INbreast datasets, achieving classification accuracies of 95.95%, 99.39%, and 96.53%, respectively. Various preprocessing techniques enhanced data quality, such as CLAHE, cropping, and data augmentation. The study highlighted the effectiveness of CNN classifiers, especially VGG-19, in distinguishing between benign and malignant masses, demonstrating the advantages of DL over traditional methods.

Chattopadhyay et al. [154] introduced Multi-scale Dual Residual Recurrent Network (MTRRE-Net), a deep-learning model for breast cancer detection using histopathological images from the BreaKHis dataset. This model features multi-scale learning and dual residual recurrent connections to address the vanishing gradient problem. MTRRE-Net outperformed state-of-the-art methods in classification accuracy, effectively capturing fine details in the images, especially across different magnification levels, and showed efficient performance with smaller datasets.

Belhaj Soulami et al. [83] developed a novel capsule network architecture for classifying suspicious regions in digital mammograms, enhancing early breast cancer detection. Their efficient model, with fewer parameters than traditional capsule networks, was trained on diverse

mammogram datasets using data augmentation. It achieved a binary classification accuracy of 96.03% and a multi-class accuracy of 77.78%, aiding radiologists in categorizing breast masses as normal, benign, or malignant. This advancement could significantly minimize unnecessary clinical procedures and biopsies.

Lu et al. [115] developed Spatial Attention Fusion Network (SAFNet), a model combining deep learning and classifier fusion for enhanced breast cancer detection and lesion localization in ultrasound images. SAFNet employs a pre-trained ResNet-18 architecture, enhanced with a spatial attention module for feature extraction. These extracted features were classified using three shallow neural networks: ELM, Random Vector Functional Link (RVFL), and Spiking Neural Network (SNN), with the final classification determined by majority voting. Evaluated on a public breast ultrasound dataset, SAFNet achieved a high classification accuracy of 94.10%, outperforming existing methods such as CNN-GTD and FR-CNN.

Liu et al. [142] utilized a hybrid deep learning approach with The Cancer Genome Atlas-Breast Invasive Carcinoma (TCGA-BRCA) dataset, incorporating gene expression data, copy number variation (CNV), and pathological images from 831 breast cancer patients. After Z-score standardization and PCA, 398 gene features were analyzed by a five-layer DNN, achieving 85.06% accuracy. Meanwhile, pathological images were processed and analyzed using a 28-layer CNN with VGG16, attaining 72.77% accuracy. A multimodal fusion model combining both methods outperformed the individual models, achieving 88.07% accuracy and an AUC of 0.9427. This study demonstrates the effectiveness of multimodal deep learning, enhancing breast cancer diagnosis and prognosis.

Kashyap et al. [151] introduced the Dilated Residual Grooming Kernel (DRGK) model for breast cancer detection, utilizing the BreakHis and BreCaHAD datasets. They implemented stain normalization and data augmentation with 19 parameters to address color divergence and prevent overfitting. It incorporates various convolutional units and achieved impressive accuracies of 97.37% on BreakHis and 97.50% on BreCaHAD, outperforming VGG16, VGG19, and ResNet50, demonstrating significant enhancements in breast cancer detection accuracy.

Lu et al. [116] introduced Breast Cancer Detection Network (BCDNet), a CAD system for breast cancer detection in ultrasound images using deep learning. BCDNet employs TL with CNN models, particularly identifying ResNet-50 as the optimal backbone, integrated with an ELM classifier, optimized with a Bat Algorithm and Chaotic Maps (BACM). The results show that ResNet-50 achieves the best accuracy of 93.97%, outperforming other architectures. The BACM optimization further enhances performance, establishing BCDNet as a leading CAD system for breast cancer detection in ultrasound images.

Asadi et al. [104] proposed an efficient cascaded network architecture for breast cancer diagnosis, combining UNet for tumor segmentation and a simplified, eight-layer ResNet50-based network for classification. The classification stage yielded a high accuracy of 98.61% and an F1-score of 98.41%, surpassing other models. Further comparative evaluations validated the model's effectiveness in breast cancer diagnosis, supported by comprehensive experimental results.

Butun et al. [182] investigated automated cancer metastasis detection in lymph nodes using the PCAM dataset of 220,025 histopathology images. They evaluated various ResNet architectures (ResNet-34, ResNet-50, ResNet-101) and found that ResNet-101, trained without pre-training, achieved the best accuracy of 98.60%. Integrating Convolutional Block Attention Modules (CBAM) with ResNet-50 improved performance to 98.58%. This approach shows promising performance in elevating the diagnostic burden of breast cancer detection.

Wang et al. [117] introduced the Multitask Information Bottleneck Network (MIB-Net) for breast tumor diagnosis, which maximizes mutual information between latent characteristics and labels while reducing the impact of the input data, thus improving interpretability. It includes a Contribution Score Map (CSM) to visualize decision making and use multitask learning with dual prior guidance to improve task correlation. MIB-Net outperformed existing methods for breast cancer classification.

Gerbasi et al. [93] presented Deep MIcroCAlcifications (DeepMiCa), an automated DL pipeline to analyze mammograms containing microcalcifications. The pipeline encompasses several steps, including preprocessing, tumor segmentation utilizing a UNet architecture, and lesion classification through transfer learning with VGG16 and ResNet18, leveraging the INbreast and CBIS-DDSM datasets. DeepMiCa is optimized to require less computational power and provides visual explanations through Grad-CAM and SHapley Additive exPlanations (SHAP) methods. It achieves an impressive AUC of 0.89 for classification, underscoring its value as a tool for radiologists.

Shamshiri et al. [149] proposed a TL framework designed for breast cancer classification in scenarios with limited annotated data. To bridge the domain gap between natural and medical images, they pre-trained CNNs on histopathological images. The study used six CNN architectures for segmentation and classification to compare three training scenarios—pre-training on histopathological images, pre-training on ImageNet, and training from scratch. The method using histopathological images achieved a validation accuracy of 98.73%, surpassing previous methods and demonstrating its effectiveness for breast cancer classification.

Aslan et al. [105] classified mammography images as normal, benign, or malignant using the MIAS (322 images) and INbreast (115 cases) datasets. After preprocessing steps like thresh-

olding and data augmentation, they developed two deep learning models: a CNN and a hybrid CNN with BiLSTM. The hybrid model achieved accuracies of 98.56% on MIAS and 92.26% on INbreast, outperforming the individual CNNs, highlighting the effectiveness of combining preprocessing techniques with a hybrid deep learning architecture for high diagnostic accuracy in mammography classification.

Yurdusev et al. [82] improved microcalcifications visibility in mammograms using a difference filter before classifying with deep learning models. They tested Faster R-CNN and YOLOv4 on 500 images from the DDSM database. The difference filter enhanced MC prominence, and data augmentation expanded the dataset. YOLOv4 initially achieved 91.33% accuracy, while Faster R-CNN reached 88.67%. Notably, applying the difference filter boosted YOLOv4's accuracy to 97.67%, reducing false detections and improving sensitivity, showcasing the difference filter's impact on deep learning models for breast cancer detection.

Deb and Jha [118] proposed a fuzzy-rank-based ensemble network for classifying breast ultrasound images into benign, malignant, and normal categories. They fine-tuned four pre-trained CNNs—VGG-Net, DenseNet, Xception, and Inception—on a dataset of 780 images from Kaggle. Using a fuzzy-rank method to integrate predictions, the ensemble model achieved an accuracy of 85.23±2.52% with five-fold cross-validation, outperforming individual models. This approach effectively enhances breast cancer detection by combining the strengths of multiple classifiers.

Table 2.20: Performance comparison of deep learning techniques for breast cancer diagnosis using imaging datasets (Part1).

Refs	**Year**	**Dataset(s)**	**Method(s)**	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**Precision (%)**	**F1-score (%)**	**AUC (%)**
[155]	2019	BACH, BreakHis	EMS-Net	99.75	-	-	-	-	99.99
[94]	2020	CBIS-DDSM, INBreast, MIAS	VGG-FTED	-	-	-	-	93.5	89.5
[81]	2020	DDSM, INbreast	Modified InceptionResNet-V2	97.50	-	-	-	-	-
[147]	2021	BreaKHis	PTGAN	88.9	87.1	-	89.0	88.0	85.8
[153]	2021	BreaKHis	ResNet-18	98.42	99.01	-	98.75	98.88	-
[115]	2022	BUSI	SAFNet (with spatial attention)	94.10	94.93	-	98.14	96.50	-
[142]	2022	TCGA-BRCA	Hybrid DL model	88.07	-	-	-	-	-
[151]	2022	BreakHis, BreCaHAD	DRGK Model	96.33	97.0	-	86.0	97.0	96.89
[152]	2022	BreakHis	ResNet-18	99.7	97.53	97.8	99.59	-	-
[154]	2022	BreakHis	MTRRE-Net	97.81	94.0	-	96.0	95.0	-

Table 2.21: Performance comparison of deep learning techniques for breast cancer diagnosis using imaging datasets (Part2).

Refs	Year	Dataset(s)	Method(s)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)	AUC (%)
[106]	2022	MIAS, DDSM, INbreast	VGG-19	99.39	-	-	-	-	-
[83]	2022	DDSM, CBIS-DDSM, INbreast	CapsNet	78.81	86.33	85.16	74.95	80.12	90.6
[118]	2023	BUSI	Fuzzy ensemble-based model	85.23	-	-	-	-	-
[82]	2023	DDSM	Yolov4	91.33	93.44	89.89	-	-	-
[105]	2023	MIAS, INbreast	CNN-BiLSTM	92.26	86.21	92.95	91.22	88.53	-
[182]	2023	PCam	ResNet-101	98.60	-	-	-	-	98.53
[117]	2023	BUSI	MIB-Net	92.97	92.97	-	93.21	92.85	98.65
[93]	2023	INbreast, CBIS-DDSM	ResNet18	83.0	75.0	88.0	81.0	-	89.0
[149]	2023	BreakHis	DenseNet169, VGG-16	94.55	-	-	-	-	-
[116]	2023	BUSI	BCDNet	93.97	95.24	-	97.68	96.42	-
[104]	2023	INbreast	ResNet50	96.8	96.8	-	96.8	96.8	-

2.6 Conclusion

This chapter has provided a thorough overview of the key datasets, pre-processing techniques, and classification methods used in breast cancer research. Medical imaging datasets, such as mammograms, MRI, and histopathological images, have enabled the development of powerful ML models capable of detecting and classifying breast tumors. The integration of clinical datasets further complements imaging data, offering a more comprehensive view of breast cancer diagnosis. Pre-processing techniques have proven essential in preparing data for analysis, ensuring that features relevant to tumor detection are accurately captured and modeled. Furthermore, the comparison of traditional machine learning methods, such as SVM and RF, with more advanced deep learning architectures, including CNNs and LSTMs, highlights the progression of the field towards more automated and efficient diagnostic tools. Notably, the application of hybrid methods and the development of novel objective functions, as seen in recent literature, offer promising avenues for further research and improvement in classification performance. This review underscores the importance of continuous innovation in data utilization and algorithm development. The combination of rich datasets, advanced pre-processing, and SOTA classification methods holds the potential to significantly enhance the accuracy and reliability of breast cancer detection systems, ultimately contributing to better clinical outcomes.

Chapter 3

BREAST CANCER CLASSIFICATION USING FEATURE SELECTION, DATA BALANCING AND HYPERPARAMETERS OPTIMIZATION

3.1 Introduction

This chapter explores significant challenges in breast cancer classification, emphasizing feature selection, data imbalance, and model parameter optimization. A malignancy correlation-based technique is used to identify highly discriminative features, enhancing data representation and directing the model's focus toward the most relevant information. To address data imbalance and improve performance in minority classes, several balancing techniques, such as SMOTE, KNNOR oversampling, and RUS undersampling, are investigated. Moreover, hyperparameters optimization maximizes classification accuracy in various scenarios, ensuring optimal model performance. These three strategies are integrated into an innovative framework for breast cancer classification, leveraging their combined benefits. The effectiveness of this integrated framework is assessed using seven established ML algorithms (SVM, KNN, DT, RF, NB, LR, and XGBoost), as well as DNNs to showcase its classification capabilities. Ultimately, the framework's performance is evaluated against state-of-the-art techniques, ensuring an objective assessment of its relative effectiveness. These advancements aim to enhance the accuracy and

efficiency of BC classification, ultimately supporting healthcare professionals by streamlining decision-making processes and reducing resource requirements.

3.2 Proposed System

Breast cancer classification faces significant challenges, including identifying relevant features, addressing data imbalance issues, and optimizing model parameters. Many recent SOTA studies have addressed these issues separately, resulting in models that lack robustness and accuracy. To overcome these limitations, we propose a novel framework that integrates feature selection, data balancing, and hyperparameters optimization into a well-established workflow for breast cancer classification. We begin our approach by applying data cleaning to the benchmark datasets DS1, DS2, DS3, and DS4, which we will define later. Subsequently, a malignancy correlation approach selects the most pertinent features for each dataset. For datasets exhibiting significant class imbalance, specifically DS1, DS2, and DS3, we employed both oversampling and undersampling techniques to ensure balanced data representation. These datasets were subdivided into training and testing subsets. Then, seven ML algorithms and DNNs were implemented to perform the training process based on the training data samples and build the trained models. This comprehensive approach enables the trained models to classify testing data effectively. Our method significantly enhances the diagnostic accuracy of traditional ML algorithms and DNNs by optimizing these key aspects, outperforming existing SOTA techniques. The efficacy of our approach is substantiated by experimental results using four publicly available datasets. Figure 3.1 provides an overview of the proposed schematic workflow.

3.2.1 Data Description

Assessing different machine learning methods requires the use of multiple publicly accessible datasets. This subsection presents an in-depth overview of the four well-known breast cancer datasets employed in our study, including WDBC (DS1), WBCD (DS2), WPBC (DS3), and Coimbra (DS4), which were defined in section 2.2.1. Table 3.1 summarizes significant details such as the number of samples, attributes, and samples within each class for each dataset, and they are further detailed individually in Tables 3.2, 3.3, 3.4, and 3.5, respectively.

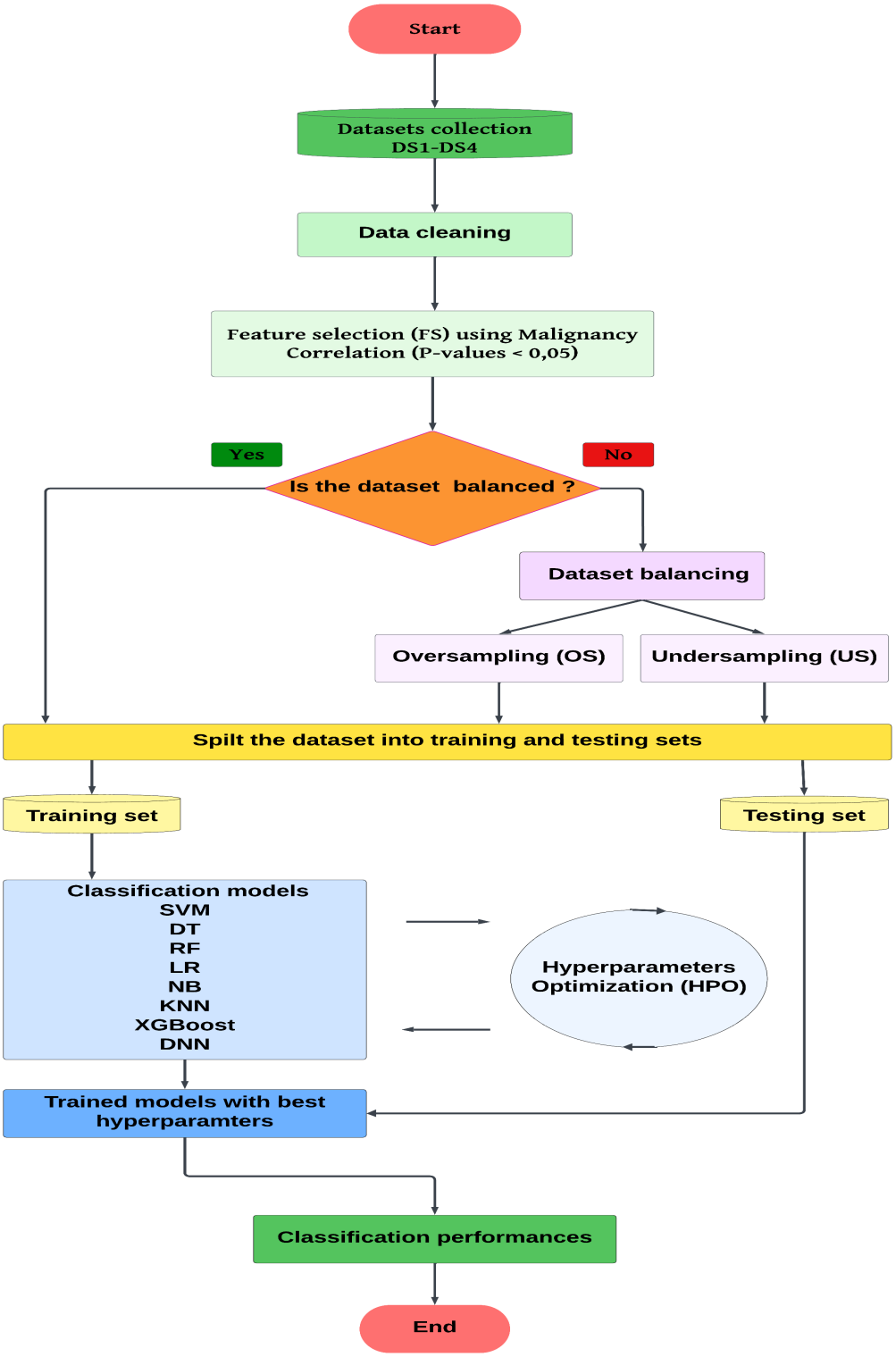


Figure 3.1: Scheme illustrates the proposed new framework, highlighting key steps such as data preprocessing, data splitting, and training-testing phases.

Table 3.1: Overview of breast cancer datasets: attributes, samples, and class distribution.

Dataset	Name	Number of Attributes	Number of Samples	Class Distribution (Balanced/Unbalanced)
DS1	Wisconsin Diagnosis Breast Cancer (WDBC)	33	569	212 Malignant, 357 Benign (Unbalanced)
DS2	Wisconsin Breast Cancer (WBCD)	10	699	241 Malignant, 458 Benign (Unbalanced)
DS3	Wisconsin Prognostic Breast Cancer (WPBC)	34	198	47 Recurrent, 151 Non-recurrent (Unbalanced)
DS4	Breast Cancer Coimbra	10	116	64 Not Healthy, 52 Healthy (Balanced)

Table 3.2: A detailed attributes description of the WDBC dataset.

No.	Attribute	Description
1	ID	The patient identification number.
2	Diagnosis	M = malignant, B = benign
3	Radius_mean	6.981 to 28.11
4	Texture_mean	9.71 to 39.28
5	Perimeter_mean	43.79 to 188.5
6	Area_mean	143.5 to 2501.0
7	Smoothness_mean	0.05263 to 0.1634
8	Compactness_mean	0.01938 to 0.3454
9	Concavity_mean	0.00 to 0.4268
10	Concave_points_mean	0.00 to 0.2012
11	Symmetry_mean	0.1060 to 0.3040
12	Fractal_dimension_mean	0.04996 to 0.09744
13	Radius_se	0.1115 to 2.873
14	Texture_se	0.3602 to 4.885
15	Perimeter_se	0.757 to 21.98
16	Area_se	6.802 to 542.2
17	Smoothness_se	0.001713 to 0.03113
18	Compactness_se	0.002252 to 0.1354
19	Concavity_se	0.0 to 0.396
20	Concave_points_se	0.00 to 0.05279
21	Symmetry_se	0.007882 to 0.07895
22	Fractal_dimension_se	0.000895 to 0.02984
23	Radius_worst	7.93 to 36.04
24	Texture_worst	12.02 to 49.54
25	Perimeter_worst	50.41 to 251.20
26	Area_worst	185.20 to 4254.00
27	Smoothness_worst	0.07117 to 0.2226
28	Compactness_worst	0.02729 to 1.0580
29	Concavity_worst	0.00 to 1.2520
30	Concave_points_worst	0.00 to 0.2910
31	Symmetry_worst	0.1565 to 0.6638
32	Fractal_dimension_worst	0.05504 to 0.2075

Table 3.3: A detailed attributes description of the WBCD dataset.

No.	Attribute	Description
1	Sample Code Number	The identification number of the sample.
2	Clump Thickness	Values ranging from 1 to 10, representing the thickness of the clumps.
3	Uniformity of Cell Size	Values ranging from 1 to 10, indicating the uniformity of cell size.
4	Uniformity of Cell Shape	Values ranging from 1 to 10, representing the uniformity of cell shape.
5	Marginal Adhesion	Values ranging from 1 to 10, representing how well the cells stick together.
6	Single Epithelial Cell Size	Values ranging from 1 to 10, indicating the size of single epithelial cells.
7	Bare Nuclei	Values ranging from 1 to 10, representing the presence of bare nuclei.
8	Bland Chromatin	Values ranging from 1 to 10, indicating the texture of the chromatin.
9	Normal Nucleoli	Values ranging from 1 to 10, representing the size and shape of nucleoli.
10	Mitoses	Values ranging from 1 to 10, indicating the frequency of mitosis.
11	Target Label	M = malignant, B = benign

Table 3.4: A detailed attributes description of the WPBC dataset.

No.	Attribute	Description
1	ID	The identification number of the sample.
2	Outcome	R = recurrent, N = non-recurrent.
3	Radius_mean	Mean of distances from center to points on the perimeter.
4	Texture_mean	Standard deviation of gray-scale values.
5	Perimeter_mean	Perimeter of the cell nucleus.
6	Area_mean	Area of the cell nucleus.
7	Smoothness_mean	Local variation in radius lengths.
8	Compactness_mean	Perimeter^2 / area - 1.0.
9	Concavity_mean	Severity of concave portions of the contour.
10	Concave_points_mean	Number of concave portions of the contour.
11	Symmetry_mean	Symmetry of the cell nucleus.
12	Fractal_dimension_mean	"Coastline approximation" - 1.
13	Radius_se	Standard error of the radius.
14	Texture_se	Standard error of the texture.
15	Perimeter_se	Standard error of the perimeter.
16	Area_se	Standard error of the area.
17	Smoothness_se	Standard error of the smoothness.
18	Compactness_se	Standard error of the compactness.
19	Concavity_se	Standard error of the concavity.
20	Concave_points_se	Standard error of the concave points.
21	Symmetry_se	Standard error of the symmetry.
22	Fractal_dimension_se	Standard error of the fractal dimension.
23	Radius_worst	Largest mean value for radius.
24	Texture_worst	Largest mean value for texture.
25	Perimeter_worst	Largest mean value for perimeter.
26	Area_worst	Largest mean value for area.
27	Smoothness_worst	Largest mean value for smoothness.
28	Compactness_worst	Largest mean value for compactness.
29	Concavity_worst	Largest mean value for concavity.
30	Concave_points_worst	Largest mean value for concave points.
31	Symmetry_worst	Largest mean value for symmetry.
32	Fractal_dimension_worst	Largest mean value for fractal dimension.
33	Tumor Size	Largest diameter of the excised tumor.
34	Lymph Node Status	Number of positive axillary lymph nodes observed.

Table 3.5: A detailed attributes description of the Coimbra dataset.

No.	Attribute	Description
1	Age (years)	Age of the patient.
2	BMI (kg/m²)	Body Mass Index of the patient.
3	Glucose (mg/dL)	Glucose level in the blood.
4	Insulin (µU/mL)	Insulin level in the blood.
5	HOMA	Homeostasis Model Assessment (insulin resistance measure).
6	Leptin (ng/mL)	Leptin hormone level.
7	Adiponectin (µg/mL)	Adiponectin hormone level.
8	Resistin (ng/mL)	Resistin hormone level.
9	MCP-1 (pg/dL)	Monocyte Chemoattractant Protein-1 level.
10	Status	Health status: 1 = healthy, 2 = diagnosed with breast cancer.

3.2.2 Data Preprocessing

Data preprocessing is essential because it converts the dataset into a practical format that can be effectively utilized by the algorithms. In this study, we employed several preprocessing strategies, which we will outline in the following sub-sections.

3.2.2.1 Data Cleaning

Before inputting data into machine learning algorithms, it is essential to check for missing values, which are typically indicated as NaN or None. Addressing these missing values is crucial for maintaining data integrity. This can be achieved by either removing attributes with missing values or filling them using an automated workflow [39]. In this study, we identified attributes with missing values across four datasets: the WDBC, WBCD, WPBC, and Coimbra datasets, as shown in Figure 3.2. Each dataset presented unique challenges in dealing with missing data, necessitating customized approaches to ensure the quality of the analysis.

- **DS1**: The attribute *"Unnamed: 32"* contained over 500 missing entries. Due to the extensive missing data, this attribute was deemed unsuitable for analysis and was removed entirely.
- **DS2**: The *"Bare_nuclei"* attribute had 16 missing values. These values were imputed using statistical methods, such as mean imputation, to preserve the dataset's completeness while minimizing data loss.
- **DS3**: The attribute *"X.34"* exhibited four missing values. These were also replaced using statistical imputation techniques to maintain the integrity and usability of the dataset.

- **DS4**: This dataset was free of missing values, requiring no imputation or attribute removal. It was ready for ML analysis without additional preprocessing.

This careful examination and treatment of missing data ensured that each dataset was appropriately prepared for subsequent analysis, minimizing the risk of bias or inaccuracies introduced by incomplete data. The strategies employed removing attributes with substantial missing data or imputing values where feasible, demonstrating the importance of a targeted approach to data cleaning in maintaining the robustness of machine learning models.

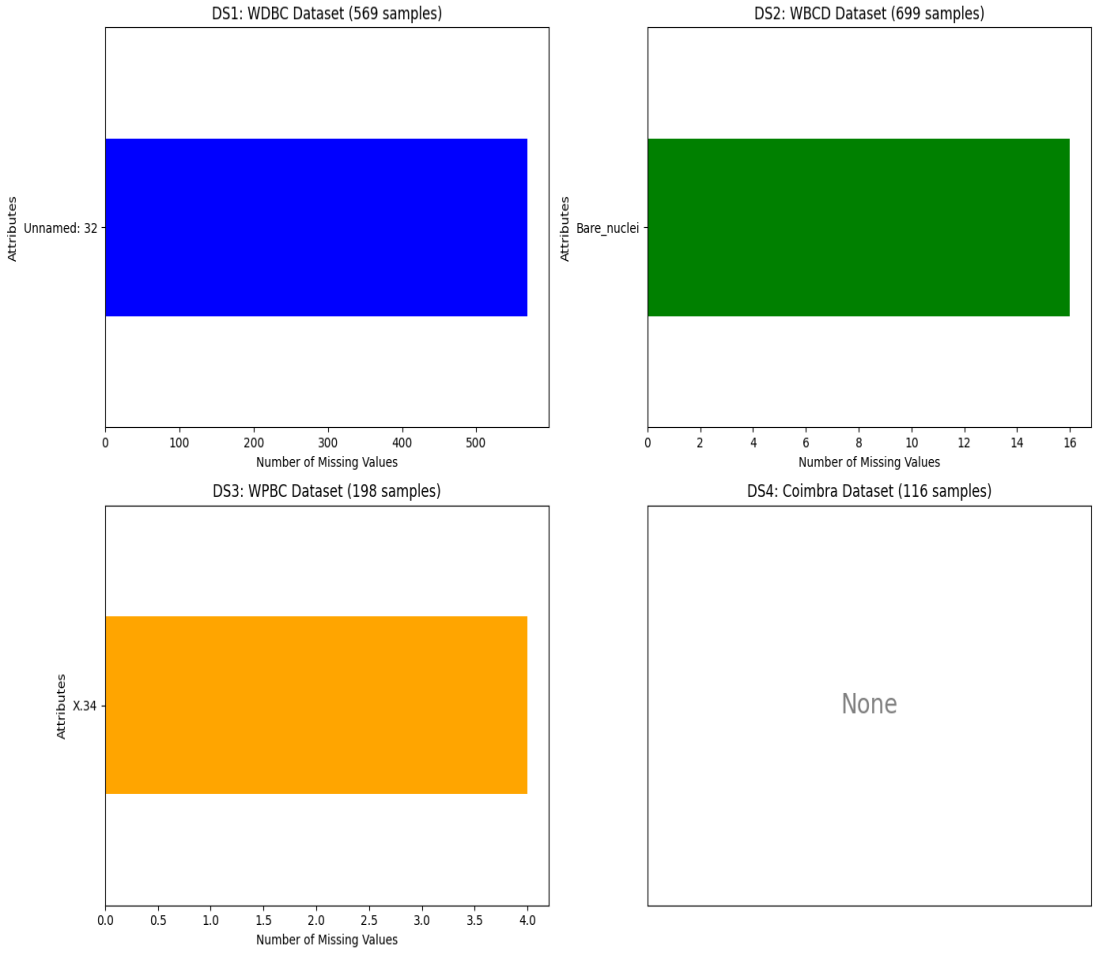


Figure 3.2: Visualization of missing values across breast cancer datasets: WDBC, WBCD, WPBC, and Coimbra.

3.2.2.2 Feature Selection

Feature selection is necessary to improve the model's performance, influencing model accuracy. In this subsection, we employed malignancy correlation analysis between the individual features and the target (label) to determine the most significant features for each dataset. Features with high correlation with the target were selected, whereas those with low correlation were dropped [39]. The selection process was used by analyzing the p-values from the Pearson Correlation Coefficient (PCC) to measure the significance of the correlations between each feature and the target variable. Features with p-values less than 0.05 were considered to have a strong correlation with the target and were selected.

A. Pearson Correlation Coefficient: The PCC (denoted as r) is a measure of the linear correlation between two variables and frequently used to identify the correlation of a feature (X) and the target variable (Y). It can be determined as shown below in Eq. (3.1):

$$r = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_i (X_i - \overline{X})^2 \sum_i (Y_i - \overline{Y})^2}} \tag{3.1}$$

Where: X_i and Y_i are the individual data points, and $\overline{X}$ and $\overline{Y}$ represent their means, respectively.

B. P-value Calculation: The p-value serves as an essential metric; it can measure the relevance of the correlation, assuming that the null hypothesis indicates no correlation between the feature and the target. The null hypothesis is ignored only when the p-value exceeds a pre-defined significance level α. It is determined on the basis of the t-test, which is formally expressed in Eq (3.2) as follows:

$$\text{p-value} = P\left(T \geq \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}}\right) \tag{3.2}$$

Where:

T is the t-statistic, $|r|$ is the absolute value of the correlation coefficient, and n is the number of data points.

The effect of the feature selection process on the datasets is illustrated in Table 3.6. It can be observed that the number of features significantly decreased after selection, helping reduce dimensionality while retaining relevant information. This demonstrates the effectiveness of feature selection in eliminating irrelevant features, thus simplifying the models and potentially enhancing their performance.

- **DS1**: The number of features was reduced from 30 to 25 after feature selection.
- **DS2**: Maintained 9 features before and after selection, indicating the original features were highly correlated with the target.
- **DS3**: The features were reduced from 32 to 10 after feature selection.
- **DS4**: The features decreased from 9 to 4 after selection.

Table 3.6: A comparison of features counts before and after feature selection across breast cancer datasets: WDBC, WBCD, WPBC, and Coimbra.

Datasets	**Total features**	**Selected features**	**The selected features**
DS1 (WDBC)	30	25	'concave points_worst', 'perimeter_worst', 'concave points_mean', 'radius_worst', 'perimeter_mean', 'area_worst', 'radius_mean', 'area_mean', 'concavity_mean', 'concavity_worst', 'compactness_mean', 'compactness_worst', 'radius_se', 'perimeter_se', 'area_se', 'texture_worst', 'smoothness_worst', 'symmetry_worst', 'texture_mean', 'concave points_se', 'smoothness_mean', 'symmetry_mean', 'fractal_dimension_worst', 'compactness_se', 'concavity_se'
DS2 (WBCD)	9	9	'Uniformity_of_cell_shape', 'Uniformity_of_cell_size', 'Bare_nuclei', 'Bland_chromatin', 'Clump_thickness', 'Normal_nucleoli', 'Marginal_adhesion', 'Single_epithelial_cell_size'
DS3 (WPBC)	32	10	'X.26', 'X.23', 'X.25', 'X.6', 'X.33', 'X.5', 'X.3', 'X.34', 'X.16', 'X.15'
DS4 (Coimbra)	9	4	'Glucose', 'HOMA', 'Insulin', 'Resistin'

3.2.2.3 Data Balancing

Data balancing is a crucial procedure that aims to achieve balance within the dataset's classes, ensuring that majority and minority classes have approximately equal samples [246]. A class

imbalance in the training dataset can potentially impact the classifier performance by boosting the accuracy of the majority class at the expense of the minority class. In this study, we used two sampling approaches, oversampling and undersampling to efficiently deal with the class-imbalance issue, which is a vital factor in training models. It is worth noticing that DS1, DS2, and DS3 have many positive samples (benign), indicating a high level of class imbalance, while DS4 has a more balanced distribution between its two classes (both classes are almost equivalent).

Oversampling: Oversampling methods are particularly effective because they increase the number of samples in the minority class, helping to overcome the problem of class imbalance. Two widely used oversampling techniques were used in this study: the synthetic minority oversampling technique and the k-nearest neighbor oversampling technique. These techniques provide artificial samples for the minority class, balancing the dataset and improving the model performance in classification tasks. Figure 3.3 illustrates the class distributions before and after oversampling for the four datasets used in this study (DS1-DS4). The majority class remained unchanged in size across all datasets, while the minority class was significantly augmented, as the following:

- For DS1 (WDBC), the number of minority samples increased from 212 to 357 after applying oversampling techniques.
- For DS2 (WBCD), the minority samples were augmented from 241 to 458, achieving a balance similar to the majority class.
- For DS3 (WPBC), the number of minority class samples rose from 47 to 151 through oversampling.
- For DS4 (Coimbra), which initially had a minor class imbalance was adjusted and treated as a balanced dataset.

These adjustments effectively mitigate bias towards the majority class, enhancing classification accuracy. The visual results validate that SMOTE and KNNOR successfully oversampled the minority class, equating it to the majority class and ensuring a more balanced dataset.

A. Synthetic Minority Over-sampling Technique: The SMOTE is an oversampling method that aims to address class imbalance issues over datasets in machine learning. Unlike Random Oversampling (ROS), which duplicates minority class samples, SMOTE generates synthetic instances within the minority class's feature space, preventing overfitting. It identifies

nearby minority class samples and creates new ones between them by choosing a random sample and its nearest neighbors, typically five by default, to produce synthetic samples [247].

B. K-Nearest Neighbor Oversampling Technique: The KNNOR is considered as a promising solution to the issue of class imbalance in classification and regression tasks by generating synthetic samples for the underrepresented minority class [248, 249]. In contrast to the SMOTE method, KNNOR addresses issues such as noisy data, tiny disjuncts, and within-class imbalances [248]. This latter comprises a revolutionary filtering approach that aids in identifying the critical and safe areas for augmentation and generating synthetic data points of the minority class, resulting in more representative data. The KNNOR's main three steps are as follows [248, 250]:

- Select a starting minority instance from the dataset as the initial point for oversampling.
- Determine K, the number of nearest neighbors of the selected minority instance to consider.
- Generate synthetic data points for the minority class based on the selected minority instance and its k-nearest neighbors, considering the relative density of the entire population.

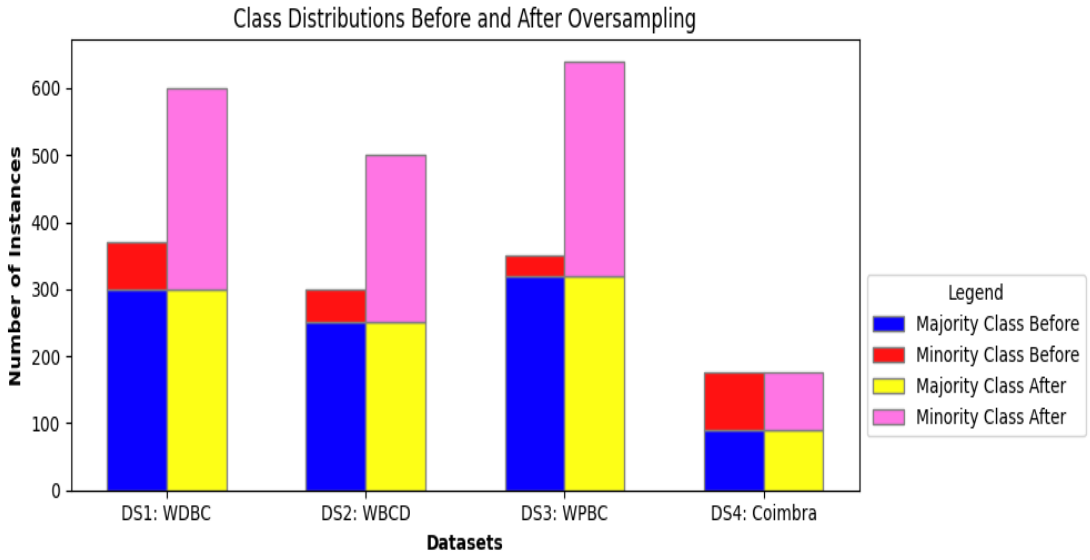


Figure 3.3: The impact of oversampling on class distribution across breast cancer datasets: WDBC, WBCD, WPBC, and Coimbra.

Undersampling: Beside oversampling techniques, undersampling techniques are used to reduce the number of instances in the majority class, creating a more balanced dataset. Figure 3.4 depicts the class distributions before and after applying RUS on the four datasets (DS1-DS4). The majority class was significantly reduced in all datasets, resulting in a more balanced class distribution.

- In DS1 (WDBC), the majority class was undersampled from 357 to approximately 212 instances.
- In DS2 (WBCD), which initially exhibited a significant imbalance, the majority class was reduced from about 500 to 241 instances.
- In DS3 (WPBC), the majority class was decreased from 151 to around 50 instances, achieving a balance with the minority class.
- In DS4 (Coimbra), both classes appeared to have an almost equal number of samples.

This rebalancing helped to mitigate the potential bias in training models and improved the classification performance across the datasets. By applying RUS, the algorithms could focus more on the minority class, improving their ability to correctly classify instances from both classes.

A. Random Undersampling: The RUS is one of the simplest undersampling methods. This approach aims to balance a dataset class's distribution by reducing the size of the majority class until a balanced class distribution is achieved. This makes it possible for machine learning algorithms to address the class-imbalanced issue without introducing synthetic instances [247, 251].

3.2.2.4 Data Normalization

Data normalization is commonly referred to as the process of data organization. The StandardScaler standardization method is used in our case. This method transforms the feature's values to have a mean of 0 and a standard deviation of 1, thereby conforming to a standard normal distribution. Eq (3.3) illustrates the formula for StandardScaler, where: X is the input data, $\overline{X}$ stands for the mean value, σ_X represents the standard deviation, and X_{SS} denotes the standardized attribute values [252].

$$X_{\text{SS}} = \frac{X - \bar{X}}{\sigma_X} \tag{3.3}$$

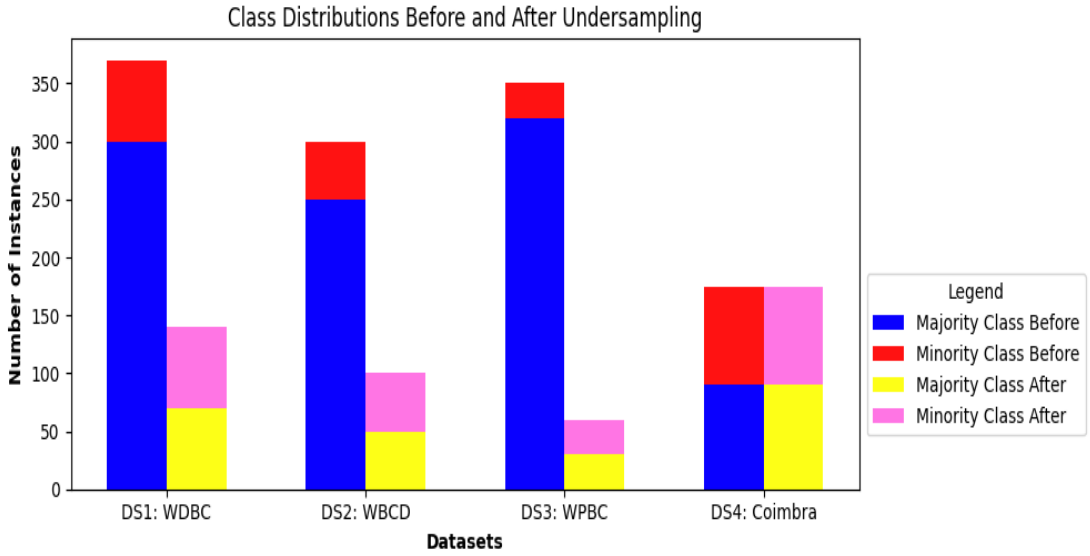


Figure 3.4: The impact of undersampling on class distribution across breast cancer datasets: WDBC, WBCD, WPBC, and Coimbra.

3.2.3 Classification Methods

We have used eight ML techniques to evaluate the proposed method. These included conventional algorithms, such as SVM, DT, RF, NB, LR, KNN, and XGBoost, in addition to DNNs, leveraging their ability to learn complex patterns. The unique characteristics of each technique are detailed in Table 3.7, underscoring their suitability for various classification tasks.

Table 3.7: Overview of the used techniques and their key features for breast cancer classification.

Technique	**Key Features**
SVM	Constructs hyperplane for class separation.
DT	Tree-structured; splits data based on feature values.
RF	Ensemble of decision trees; averages multiple models.
NB	Probabilistic, assuming feature independence.
LR	Linear model for binary classification.
KNN	Classifies based on majority class among nearest neighbors.
XGBoost	Optimized gradient boosting; combines weak learners.
DNNs	Multiple layers; learns complex patterns.

3.2.4 Hyperparameters Optimization

Hyperparameters optimization plays a crucial role in enhancing the performance of ML models. It involves a systematic search for the optimal combination of hyperparameters and fine-tuning the model to attain the highest level of performance. This process significantly impacts model accuracy and overall effectiveness, especially in complex tasks like breast cancer classification. It includes three HPO techniques: grid search, random search, and bayesian optimization [253, 254]. We selected grid search for its exhaustive exploration of the identified parameter space, ensuring that all possible combinations are evaluated to identify the optimal configuration and optimize hyperparameters across several machine learning algorithms, including SVM, DT, RF, LR, KNN, and XGBoost, with multiple hyperparameters tested for each algorithm. In contrast, a semi-automatic approach was utilized for tuning the DNNs, using hyperparameters such as learning rates, batch sizes, number of epochs, neurons per layer, and dropout rates, which were manually adjusted based on practical observations, as presented in Table 3.8. This method offers greater flexibility and allows for intuition-based adjustments, particularly in deep learning models, where optimal configurations typically require iterative experimentation.

Table 3.8: Overview of the used hyperparameters optimization techniques and different parameters combinations for ML algorithms and DNN classifier.

Methods	Parameters	Description	Different combinations	HPO approach
SVM	C	Regularization parameter.	[0.1, 1, 10]	Automatic approach (GS)
	kernel	Specifies the kernel type to be used in the algorithm.	['linear', 'poly', 'rbf']	
DT	max_depth	Maximum depth of the tree.	[10, 20, 30]	
	min_samples_leaf	The minimum number of samples required to be at a leaf node.	[1, 3, 5]	
	min_samples_split	The minimum number of samples required to split an internal node.	[2, 5, 10]	
RF	n_estimators	The number of trees in the forest.	[5, 50, 100]	
	max_features	The number of features to consider when looking for the best split.	['auto', 'sqrt', 'log2']	
	max_depth	Maximum depth of the tree.	[10, 20, 30]	
LR	C	Inverse of regularization strength.	[0.1, 1, 10]	
	solver	Algorithm to use in the optimization problem.	['lbfgs', 'liblinear']	
	penalty	The norm used in the penalization.	['l1', 'l2']	
KNN	n_neighbors	Number of neighbors to use.	[3, 5, 7]	
	weights	Weight function used in prediction.	['uniform', 'distance']	
	p	Power parameter for the Minkowski metric.	[1, 2]	
XGBoost	learning_rate	Step size shrinkage to prevent overfitting.	[0.01, 0.1, 0.2]	
	n_estimators	Number of trees.	[50, 100, 200]	
	max_depth	Maximum depth of the tree.	[3, 4, 5]	
	gamma	Minimum loss reduction required to make a further partition.	[0, 0.25, 1.0]	
DNN	learning_rates	Learning rate used to update weights.	[0.0001, 0.001, 0.01, 0.1]	Semi-automatic approach
	batch_sizes	Number of samples per gradient update.	[16, 32, 64]	
	epochs_values	Number of passes through the entire dataset.	[200]	
	neurons_per_layer	Number of neurons per hidden layer.	[(512,128,64), (512,256,128), (512,128,32)]	
	dropout_rates	Fraction of input units to drop.	[0.1, 0.2, 0.3]	

3.2.5 Performance Evaluation Metrics

Performance evaluation is crucial for assessing classification models' effectiveness, reliability, and applicability, especially in healthcare scenarios like breast cancer, where precision is paramount. This section outlines commonly used evaluation metrics, including confusion matrix, accuracy, precision, F1-score, and others. These metrics quantify a model's ability to classify instances, showing how the model performs accurately.

3.2.5.1 Confusion Matrix

The confusion matrix (CM) is a fundamental tool for summarizing the performance of a classification algorithm by comparing actual and predicted classifications [255]. Represented as an NxN grid, where N denotes the number of classes, the CM provides a detailed analysis of correct and incorrect predictions for each class. The CM represents positive and negative outcomes in binary classification (N = 2). Its key components are:

- **True Positive (TP):** instances correctly predicted as positive (e.g., correctly identifying a patient with BC).
- **True Negative (TN):** instances correctly predicted as negative (e.g., correctly identifying a healthy individual).
- **False Positive (FP):** instances incorrectly predicted as positive (Type I error, e.g., incorrectly diagnosing a healthy individual with BC).
- **False Negative (FN):** instances incorrectly predicted as negative (Type II error, e.g., failing to diagnose a patient with BC).

While TP and TN represent correct classifications, FP and FN represent misclassifications and are fundamental for calculating various performance metrics (Eqs. 3.4-3.8). Figure 3.5 shows an illustrative example of a confusion matrix.

3.2.5.2 Accuracy

Accuracy measures the correctness of a classification model's predictions, calculated as the ratio of correctly classified instances to the total. While it is easy to interpret, accuracy can be misleading, especially in imbalanced datasets where one class greatly outnumbers others. A model may achieve high accuracy by mostly predicting the majority class, failing to perform well on the minority class.

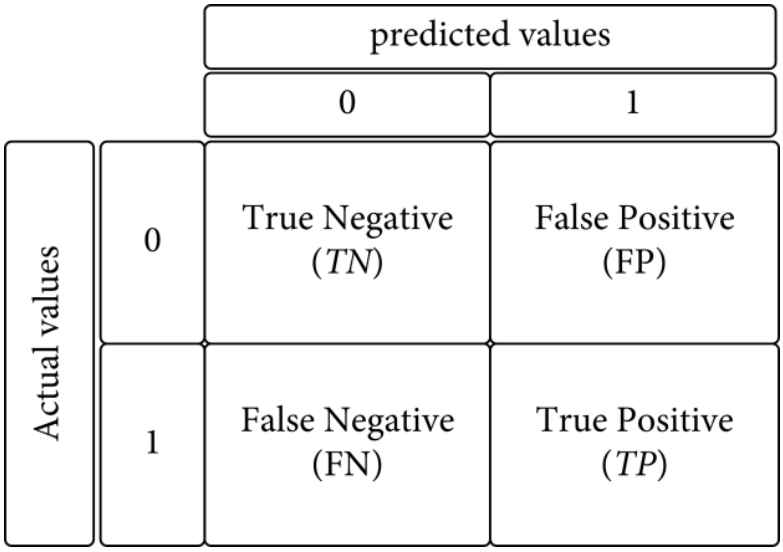


Figure 3.5: Example of a confusion matrix for binary classification.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.4}$$

3.2.5.3 Sensitivity

Sensitivity, also known as recall or the true positive rate, measures the proportion of actual positive instances (e.g., individuals with breast cancer) correctly identified by the model. This metric is particularly crucial in applications where minimizing false negatives (i.e., failing to identify individuals with the condition) is of primary importance, such as in disease screening and classification.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3.5}$$

3.2.5.4 Specificity

Specificity, also known as the True Negative Rate, measures the proportion of actual negative instances (e.g., individuals without breast cancer) correctly identified by the model. This metric is critical in scenarios where minimizing false positives (i.e., incorrectly identifying individuals as having the condition) is crucial, such as avoiding unnecessary and potentially invasive interventions for healthy individuals.

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3.6}$$

3.2.5.5 Precision

Precision, also known as the Positive Predictive Value, quantifies the proportion of instances predicted as positive (e.g., individuals predicted to have breast cancer) that are positive (i.e., indeed have the condition). This metric is particularly relevant when the cost or consequences of false positives (e.g., unnecessary anxiety, further invasive testing, or treatment) are high.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.7}$$

3.2.5.6 F1-score

The F1-score, also known as the F-measure or F-score, represents the harmonic mean of precision and sensitivity. This metric effectively balances the trade-off between these two measures and is especially valuable when assessing performance on imbalanced datasets. A perfect F1-score of 1.0 is attained only when both precision and sensitivity are perfect, meaning they are equal to 1.0.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{3.8}$$

3.3 Results and Discussion

This section examines the effectiveness of the employed ML techniques, including LR, DT, RF, SVM, NB, KNN, XGBoost, and a DNN model. These techniques have been evaluated on the four datasets (DS1, DS2, DS3, and DS4) using commonly used evaluation criteria, including accuracy, sensitivity, specificity, precision, and F1-score to accurately classify breast cancer. The selected datasets were split into training and testing sets with 80% and 20%, respectively (note: the highest performance rates were highlighted in **bold** font).

3.3.1 A Comprehensive Analysis of Feature Selection

After data cleaning, the FS is implemented to analyze all datasets (DS1, DS2, DS3, and DS4), and the performance of the used techniques is evaluated without and with feature selection, as illustrated in Table 3.9 and Table 3.10, respectively.

3.3.1.1 Without Feature Selection (WOFS)

A comprehensive analysis of the performance of various machine learning models across four distinct datasets (DS1-DS4) without implementing feature selection is provided in Table 3.9. The KNN technique stood out as the best performer for DS2, achieving an impressive accuracy of 98.57%, along with other metrics, including sensitivity, specificity, precision, and F1-score. In contrast, the DNN technique demonstrated superior performance across the remaining datasets (DS1, DS3, and DS4). Specifically, in DS1, DNN attained the highest accuracy of 98.25% and excelled in all other metrics. Similarly, for DS4, DNN recorded the best performance with an accuracy of 95.83%, showcasing high sensitivity and perfect specificity. The performance on DS3 was more varied, with no single model consistently outperforming in all metrics; however, DNN still achieved the highest accuracy despite a lower sensitivity. Overall, DNN exhibited remarkable robustness across multiple datasets, particularly in tasks requiring high classification accuracy, while KNN's strong performance in DS2 underscores its effectiveness under specific conditions.

3.3.1.2 With Feature Selection (WFS)

The effect of feature selection on the performance of various machine learning models across four distinct datasets (DS1-DS4) is illustrated in Table 3.10. In this study, low-correlation features were systematically eliminated to concentrate on the most pertinent attributes, which resulted in significant enhancements in performance for the majority of models across the datasets. Notably, the performance of models applied to DS2 exhibited minimal variation post-FS, highlighting the essential role of all features within this dataset. Conversely, dropping non-informative features contributed to improved generalization and enhanced performance in most models within the other datasets. For instance, in DS1, the DNN model distinguished itself with an accuracy of 97.36%, showing improvement over its prior performance without using FS. This enhancement was consistent across multiple performance metrics, including precision and F1-score, thereby affirming the efficacy of FS in mitigating noise and capturing crucial features.

Similarly, the SVM classifier achieved the highest performance in DS3 post-FS, reaching an accuracy of 82.5%. However, this performance was attended by variability in sensitivity and precision within the other datasets, indicating that while FS substantially benefited SVM's performance, its sensitivity to the complexities of the dataset remained a key influencing factor. In contrast, the responses of models in DS4 to FS were more heterogeneous, where the DNN and SVM techniques demonstrated comparable performance, each attaining an accuracy of 91.66%, mirroring results obtained without FS. Notably, the KNN algorithm achieved its peak perfor-

mance in DS4 post-FS, exhibiting a perfect sensitivity of 100%, underscoring the substantial advantages conferred by the FS process for this particular technique.

Table 3.9: Performance assessment of various models without feature selection (WOFS) for DS1-DS4 datasets.

Datasets	**Methods**	**Performance (%)**				
		Accuracy	**Sensitivity**	**Specificity**	**Precision**	**F1-score**
DS1	LR	96.49	93.02	98.59	97.56	95.23
	DT	93.85	90.69	95.77	92.85	91.76
	RF	96.49	93.02	98.59	97.56	95.23
	SVM	94.73	86.04	**100.0**	**100.0**	92.49
	NB	97.36	93.02	**100.0**	**100.0**	96.38
	KNN	95.61	88.37	**100.0**	**100.0**	93.82
	XGBoost	95.61	93.02	97.18	95.23	94.11
	DNN	**98.25**	**95.35**	**100.0**	**100.0**	**97.62**
DS2	LR	95.71	88.88	98.94	97.56	93.02
	DT	95.00	88.88	97.89	95.23	91.95
	RF	96.42	91.11	98.94	97.61	94.25
	SVM	96.42	93.33	97.89	95.45	94.38
	NB	96.42	97.77	95.78	91.66	94.62
	KNN	**98.57**	**97.77**	**98.94**	**97.77**	**97.77**
	XGBoost	95.00	86.66	**98.94**	97.50	91.76
	DNN	97.86	95.56	**98.95**	97.73	96.63
DS3	LR	67.50	25.00	78.12	22.22	23.52
	DT	60.00	25.00	68.75	16.66	20.00
	RF	80.00	25.00	93.75	50.00	33.33
	SVM	80.00	0.0	**100.0**	0.0	0.0
	NB	60.00	**37.50**	65.62	21.42	27.27
	KNN	75.00	50.00	81.25	40.00	**44.44**
	XGBoost	72.50	37.50	81.25	33.33	35.29
	DNN	**82.50**	12.50	**100.0**	**100.0**	22.22
DS4	LR	87.50	83.33	91.66	90.90	86.95
	DT	75.00	75.00	75.00	75.00	75.00
	RF	87.50	**100.0**	75.00	80.00	88.88
	SVM	91.66	91.66	91.66	91.66	91.66
	NB	75.00	58.33	91.66	87.50	70.00
	KNN	83.33	75.00	91.66	90.00	81.81
	XGBoost	87.50	83.33	91.66	90.90	86.95
	DNN	**95.83**	**91.67**	**100.0**	**100.0**	**95.65**

Table 3.10: Performance assessment of various models with feature selection (WFS) for the DS1-DS4 datasets.

Dataset	Methods	Performance (%)				
		Accuracy	**Sensitivity**	**Specificity**	**Precision**	**F1-score**
DS1	LR	96.49	93.02	98.59	97.56	95.23
	DT	91.22	90.69	91.54	86.66	88.63
	RF	96.49	93.02	98.59	97.56	95.23
	SVM	94.73	86.04	**100.0**	**100.0**	92.49
	NB	97.36	93.02	**100.0**	**100.0**	96.38
	KNN	95.61	88.37	**100.0**	**100.0**	93.82
	XGBoost	96.49	93.02	98.59	97.56	95.23
	DNN	**97.36**	**95.35**	**98.59**	**97.62**	**96.47**
DS2	LR	95.71	88.88	98.94	97.56	93.02
	DT	95.00	88.88	97.89	95.23	91.95
	RF	96.42	91.11	98.94	97.61	94.25
	SVM	96.42	93.33	97.89	95.45	94.38
	NB	96.42	97.77	95.78	91.66	94.62
	KNN	**98.57**	**97.77**	**98.94**	**97.77**	**97.77**
	XGBoost	95.00	86.66	98.94	97.50	91.76
	DNN	97.86	95.56	**98.95**	97.73	96.63
DS3	LR	70.00	25.00	81.25	25.00	25.00
	DT	57.50	25.00	65.62	15.38	19.04
	RF	67.50	25.00	78.12	22.22	23.52
	SVM	**82.50**	12.50	**100.0**	**100.0**	22.22
	NB	55.00	**37.50**	59.37	18.75	**25.00**
	KNN	75.00	50.00	81.25	40.00	44.44
	XGBoost	62.50	37.50	68.75	23.07	28.57
	DNN	72.50	25.00	84.38	28.57	26.67
DS4	LR	83.33	75.00	91.66	90.00	81.81
	DT	62.50	58.33	66.66	63.63	60.86
	RF	79.16	83.33	75.00	76.92	80.00
	SVM	**91.66**	91.66	**91.66**	**91.66**	**91.66**
	NB	70.83	50.00	91.66	85.71	63.15
	KNN	91.66	**100.0**	83.33	85.71	**92.30**
	XGBoost	75.00	75.00	75.00	75.00	75.00
	DNN	**91.66**	**91.66**	**91.66**	**91.66**	**91.66**

3.3.2 Data Balancing

This section provides a comprehensive analysis of the performance results obtained from testing the DS1, DS2, and DS3 datasets, using feature selection (WFS) followed by data balancing. Notably, the DS4 dataset is excluded from Tables 3.12 and 3.13, as it is inherently balanced

and does not require additional preprocessing steps.

3.3.2.1 Oversampling Techniques

In this subsection, seven machine learning algorithms—LR, DT, RF, SVM, NB, KNN, XGBoost—along with a DNN classifier, were employed to assess the performance of two oversampling techniques: synthetic minority over-sampling technique and k-nearest neighbor oversampling technique across three distinct datasets (DS1, DS2, and DS3). Table 3.12 presents the results, indicating a consistent trend where the DNN classifier outperformed all other machine learning methods across the three datasets, achieving the highest overall performance.

For DS1, models trained using SMOTE yielded results comparable to those trained with KNNOR. Notably, the DNN classifier attained an impressive accuracy of 99.30%, with perfect sensitivity (100.0%) and an F1-score of 99.33%, significantly surpassing the performance of the other models. Conversely, KNNOR demonstrated highly competitive results for both DS2 and DS3. The DNN classifier reached an accuracy of 98.91% with perfect sensitivity (100.0%) and an almost perfect F1-score of 98.91% for DS2, highlighting KNNOR's effectiveness for this particular dataset. The trend continued for DS3, where the DNN model utilizing KNNOR achieved a notable accuracy of 85.25% with an F1-score of 85.71%, once again outperforming models trained with SMOTE.

Overall, these findings emphasize the variability in the effectiveness of oversampling techniques across different datasets. While SMOTE excelled in DS1, KNNOR proved to be more effective for DS2 and DS3, illustrating the importance of selecting an oversampling technique tailored to the specific characteristics of each dataset.

3.3.2.2 Undersampling Techniques

The performance of seven ML techniques and a deep neural network model on datasets DS1, DS2, and DS3, balanced using random undersampling, is detailed in Table 3.13. The results indicate that the DNN model consistently achieved superior performance on DS1 and DS3, reaching high accuracies of 98.82% and 89.47%, respectively, along with near-perfect precision and specificity. This consistent performance underscores the robustness of the DNN and its ability to effectively handle undersampled datasets. In contrast, the performance of all used techniques is remarkably consistent for DS2. Most methods achieved an identical accuracy of 98.96%, accompanied by similar values across other performance metrics. This uniformity suggests that DS2 has characteristics that enable a wide range of models to perform equally well. The only notable exception is the DT method, which, while delivering commendable results,

shows relatively lower performance metrics than its peers.

In conclusion, the DNN model demonstrates superior effectiveness for datasets like DS1 and DS3, solidifying its suitability for handling complex and imbalanced data. On the other hand, DS2 presents a unique case where most machine learning techniques exhibit comparable performance, emphasizing the importance of tailoring model selection to the specific properties of the dataset to achieve optimal results.

3.3.3 The Impact of Hyperparameters Optimization

To comprehensively assess the influence of the hyperparameters optimization process on the performance of the chosen machine learning classifiers, we conducted a detailed analysis across four distinct datasets: DS1, DS2, DS3, and DS4. This examination aimed to prove whether the application of HPO substantially improves the predictive accuracy and reliability of these models.

For DS1, we used a combined approach that included feature selection, data balancing through the SMOTE technique, and HPO. This method will be referred to as (WFS + SMOTE + HPO). Table 3.11 provides a detailed overview of the performance metrics for each classifier under these conditions. The DNN model delivered exceptional results, achieving an impressive accuracy of 99.3%. Other classifiers also performed well, with the SVM and LR models reaching an accuracy of 98.6%.

The implementation of HPO led to significant improvements in precision and F1-scores across all models, underscoring the effectiveness of these techniques in enhancing model accuracy and reducing overfitting, particularly when combined with WFS and SMOTE.

Table 3.11: Comparison of the obtained performance of all the employed techniques with (WFS+ SMOTE+HPO) for the DS1 dataset.

Preprocessing	**Datasets**	**Methods**	**Performance (%)**				
			Accuracy	**Sensitivity**	**Specificity**	**Precision**	**F1-score**
WFS+ SMOTE	DS1	LR	98.60	98.64	98.55	98.64	98.64
		DT	94.40	94.59	94.20	94.59	94.59
		RF	96.50	95.94	97.10	97.26	96.59
		SVM	98.60	97.29	**100.0**	**100.0**	98.63
		NB	93.70	90.54	97.10	97.10	93.70
		KNN	96.50	**100.0**	92.75	93.67	96.73
		XGBoost	97.20	98.64	95.65	96.05	97.33
		DNN	**99.30**	**100.0**	98.55	98.67	**99.33**

Table 3.12: Performance assessment of various models with data balancing using the SMOTE and KNNOR techniques for the DS1-DS3 datasets.

Balancing technique	**Dataset**	**Methods**	**Performance (%)**				
			Accuracy	**Sensitivity**	**Specificity**	**Precision**	**F1-score**
SMOTE	DS1	LR	95.10	90.54	**100.0**	**100.0**	95.03
		DT	95.10	94.59	95.65	95.89	95.23
		RF	97.90	98.64	97.10	97.33	97.98
		SVM	88.81	83.78	94.20	93.93	88.57
		NB	94.40	89.18	**100.0**	**100.0**	94.28
		KNN	91.60	87.83	95.65	95.58	91.54
		XGBoost	97.90	98.64	97.10	97.33	97.98
		DNN	**99.30**	**100.0**	98.55	98.67	**99.33**
	DS2	LR	96.73	96.70	96.70	96.70	96.77
		DT	96.19	97.80	94.62	94.68	96.21
		RF	**98.36**	**100.0**	96.77	96.80	**98.37**
		SVM	**98.36**	**100.0**	96.77	96.80	**98.37**
		NB	97.28	97.80	96.77	96.73	97.26
		KNN	96.73	96.70	96.77	96.70	96.70
		XGBoost	95.65	95.60	95.69	95.60	95.60
		DNN	**98.36**	98.90	**97.85**	**97.83**	**98.36**
	DS3	LR	54.09	38.70	70.00	57.14	46.15
		DT	72.13	83.87	60.00	68.42	75.36
		RF	81.96	83.87	80.00	**81.25**	82.53
		SVM	54.09	35.48	73.33	57.89	44.00
		NB	49.18	32.25	66.66	50.00	39.21
		KNN	75.40	83.87	66.66	72.22	77.60
		XGBoost	77.00	83.87	70.00	74.28	78.78
		DNN	**83.61**	**90.32**	**76.67**	80.00	**84.85**
KNNOR	DS1	LR	93.44	88.11	**98.90**	**98.93**	93.42
		DT	92.18	91.89	92.47	92.53	92.18
		RF	95.62	97.29	93.94	93.96	95.64
		SVM	87.50	80.27	94.74	94.82	87.34
		NB	91.25	84.59	98.10	98.12	91.11
		KNN	89.37	85.14	93.60	93.63	89.32
		XGBoost	96.87	98.64	95.09	95.13	96.89
		DNN	**97.81**	**100.0**	95.71	95.76	**97.83**
	DS2	LR	94.92	94.87	94.97	94.91	94.91
		DT	95.92	96.75	95.09	95.13	95.93
		RF	97.97	98.63	97.28	97.33	97.98
		SVM	**98.98**	**99.31**	98.66	98.69	**98.98**
		NB	96.86	97.39	96.34	96.37	96.87
		KNN	95.68	95.64	95.72	95.67	95.67
		XGBoost	96.90	97.39	96.42	96.48	96.91
		DNN	**98.98**	99.17	**98.80**	**98.83**	**98.98**
	DS3	LR	53.57	40.32	66.67	53.85	46.97
		DT	70.83	80.64	60.00	66.67	73.68
		RF	80.36	83.87	**76.67**	**78.57**	81.67
		SVM	51.79	32.25	70.00	50.00	39.51
		NB	46.43	25.80	66.67	46.15	33.90
		KNN	72.32	83.87	60.00	67.57	74.77
		XGBoost	74.42	80.64	67.33	70.42	76.36
		DNN	**84.64**	**93.54**	75.56	78.45	**85.12**

Table 3.13: Performance assessment of various models with data balancing using the random undersampling technique for the DS1-DS3 datasets.

Dataset	Methods	Performance (%)				
		Accuracy	**Sensitivity**	**Specificity**	**Precision**	**F1-score**
DS1	LR	95.29	92.30	97.82	97.29	94.73
	DT	92.94	97.43	89.13	88.37	92.68
	RF	97.64	**100.0**	95.65	95.12	97.50
	SVM	91.76	87.17	95.65	94.44	90.66
	NB	94.11	92.30	95.65	94.73	93.50
	KNN	90.58	87.17	93.47	91.89	89.47
	XGBoost	97.64	97.43	97.82	97.43	97.43
	DNN	**98.82**	97.44	**100.0**	**100.0**	**98.70**
DS2	**LR**	**98.96**	**97.87**	**100.0**	**100.0**	**98.92**
	DT	97.93	95.74	**100.0**	**100.0**	97.82
	RF	**98.96**	**97.87**	**100.0**	**100.0**	**98.92**
	SVM	**98.96**	**97.87**	**100.0**	**100.0**	**98.92**
	NB	**98.96**	**97.87**	**100.0**	**100.0**	**98.92**
	KNN	**98.96**	**97.87**	**100.0**	**100.0**	**98.92**
	XGBoost	**98.96**	**97.87**	**100.0**	**100.0**	**98.92**
	DNN	**98.96**	**97.87**	**100.0**	**100.0**	**98.92**
DS3	LR	73.68	71.42	75.00	62.50	66.66
	DT	63.15	57.14	66.66	50.00	53.33
	RF	78.94	**100.0**	66.66	63.63	77.77
	SVM	63.15	57.14	66.66	50.00	53.33
	NB	68.42	57.14	75.00	57.14	57.14
	KNN	73.68	57.14	83.33	66.66	61.53
	XGBoost	57.89	57.14	58.33	44.44	50.00
	DNN	**89.47**	85.71	**91.67**	**85.71**	**85.71**

For DS2 and DS3, we applied feature selection and data balancing using the RUS technique with HPO, abbreviated as (WFS+RUS+HPO). As shown in Table 3.14, the DNN model again led the way, particularly for DS3, where it achieved an accuracy of 89.47%, significantly outperforming other models. The consistency of high performance across different datasets with varied characteristics demonstrates the robustness of the DNN model when optimized using HPO. However, it is worth noting that for DS2, nearly all classifiers, including RF, SVM, and NB, performed equally well, each achieving an accuracy close to 98.97%, indicating that the dataset's nature may allow multiple models to achieve near-optimal performance when optimized.

In the evaluation of the DS4 dataset, which was balanced without applying feature selection (referred to as WOFS), hyperparameters optimization was independently implemented, forming the approach labeled as WOFS + HPO. The results, summarized in Table 3.15, reveal that the DNN classifier consistently outperformed all other models, achieving an impressive accuracy of 95.83%. Notably, XGBoost and LR also demonstrated marked improvements, underscoring the significant impact of HPO on model performance. These findings highlight that even in the absence of feature selection, HPO remains a powerful tool for enhancing predictive accuracy and model robustness. This insight is particularly valuable for scenarios where FS is insufficient or yields limited benefits, offering a practical and effective alternative for optimizing classification performance.

Table 3.14: Comparison of the obtained performance of all the employed techniques with (WFS+RUS+HPO) for the DS2, and DS3 datasets.

Preprocessing	**Datasets**	**Methods**	**Performance (%)**				
			Accuracy	**Sensitivity**	**Specificity**	**Precision**	**F1-score**
WFS+RUS	DS2	LR	96.90	93.61	100.0	100.0	96.70
		DT	94.84	91.48	98.00	97.72	94.50
		RF	**98.97**	**97.87**	**100.0**	**100.0**	**98.92**
		SVM	97.93	**97.87**	98.00	97.87	97.87
		NB	**98.97**	**97.87**	**100.0**	**100.0**	**98.92**
		KNN	**98.97**	**97.87**	**100.0**	**100.0**	**98.92**
		XGBoost	**98.97**	**97.87**	**100.0**	**100.0**	**98.92**
		DNN	**98.97**	**97.87**	**100.0**	**100.0**	**98.92**
	DS3	LR	68.42	71.42	66.66	55.55	62.50
		DT	52.63	57.14	50.00	40.00	47.05
		RF	73.68	**85.71**	66.66	60.00	70.58
		SVM	78.94	71.42	83.33	71.42	71.42
		NB	68.42	57.14	75.00	57.14	57.14
		KNN	73.68	**85.71**	66.66	60.00	70.58
		XGBoost	73.68	71.42	75.00	62.50	66.66
		DNN	**89.47**	**85.71**	**91.67**	**85.71**	**85.71**

Table 3.15: Comparison of the obtained performance of all the employed techniques with (WOFS+HPO) for the DS4 dataset.

Preprocessing	Datasets	Methods	Performance (%)				
			Accuracy	Sensitivity	Specificity	Precision	F1-score
WOFS	DS4	LR	87.50	83.33	91.66	90.90	86.95
		DT	70.83	75.00	66.66	69.23	71.99
		RF	83.33	75.00	91.66	90.00	81.81
		SVM	83.33	75.00	91.66	90.00	81.81
		NB	75.00	58.33	91.66	87.50	70.00
		KNN	75.00	66.66	83.33	80.00	72.72
		XGBoost	87.50	91.66	83.33	84.61	87.99
		DNN	**95.83**	**91.67**	**100.0**	**100.0**	**95.65**

Overall, the HPO technique consistently improved the performance of most classification models across all datasets, especially for DNN, which surpassed all other machine learning techniques in most cases. However, it was observed that for DS2, the improvement was uniform across most techniques, indicating that while HPO is beneficial, the dataset's inherent characteristics play a crucial role in determining the overall model performance. This comprehensive assessment underscores the critical importance of HPO in the machine learning pipeline, as it can lead to significant classification performance gains across diverse datasets.

3.3.4 Comparative Analysis

To facilitate a comparative analysis of the performance of various methods, we present a summary of our experimental findings for different classifiers across all datasets (DS1, DS2, DS3, and DS4). This is accomplished using a pairwise comparison method inspired by the image quality assessment protocol based on the Accumulated Preference Index (AP_i) [256]. For two methods, A and B, with their respective accuracy values denoted as α and β, the preference index (P_i) is calculated using the decision rule outlined in Equation (3.9).

$$P_i = \begin{cases} 1, & \text{if } \alpha > \beta, \\ 0, & \text{if } \alpha < \beta, \\ 0.5, & \text{otherwise.} \end{cases} \tag{3.9}$$

The accumulated preference index for a given method is calculated by summing the P_i values across various pairwise comparisons with all other methods, as demonstrated in Tables 3.16–3.19. These tables indicate that the DNN classifier consistently achieves the highest AP_i scores

for datasets DS1, DS3, and DS4. In contrast, for dataset DS2, the DNN classifier performs on par with four other classifiers (RF, NB, KNN, and XGBoost), all achieving the same AP_i as illustrated in Table 3.17. Additionally, the DT classifier records the lowest AP_i across all datasets, highlighting its inadequacy in breast cancer classification. Overall, the DNN classifier's superior performance can be attributed to its effective integration of feature selection, data balancing, and hyperparameters optimization simultaneously.

Table 3.16: Performance comparison of different classifiers for DS1 using the accumulated preference index.

Classifier	**LR**	**DT**	**RF**	**SVM**	**NB**	**KNN**	**XGBoost**	**DNN**	AP_i
LR	–	1	1	0.5	1	1	1	0	5.5
DT	0	–	0	0	1	0	0	0	1
RF	0	1	–	0	1	0.5	0	0	2.5
SVM	0.5	1	1	–	1	1	1	0	5.5
NB	0	0	0	0	–	0	0	0	0
KNN	0	1	0.5	0	1	–	0	0	2.5
XGBoost	0	1	1	0	1	1	–	0	4
DNN	1	1	1	1	1	1	1	–	7

Table 3.17: Performance comparison of different classifiers for DS2 using the accumulated preference index.

Classifier	**LR**	**DT**	**RF**	**SVM**	**NB**	**KNN**	**XGBoost**	**DNN**	AP_i
LR	–	1	0	0	0	0	0	0	1
DT	0	–	0	0	0	0	0	0	0
RF	1	1	–	1	0.5	0.5	0.5	0.5	5
SVM	1	1	0	–	0	0	0	0	2
NB	1	1	0.5	1	–	0.5	0.5	0.5	5
KNN	1	1	0.5	1	0.5	–	0.5	0.5	5
XGBoost	1	1	0.5	1	0.5	0.5	–	0.5	5
DNN	1	1	0.5	1	0.5	0.5	0.5	–	5

Table 3.18: Performance comparison of different classifiers for DS3 using the accumulated preference index.

Classifier	**LR**	**DT**	**RF**	**SVM**	**NB**	**KNN**	**XGBoost**	**DNN**	AP_i
LR	–	1	0	0	0.5	0	0	0	1.5
DT	0	–	0	0	0	0	0	0	0
RF	1	1	–	0	1	0.5	0.5	0	4
SVM	1	1	1	–	1	1	1	0	6
NB	0.5	1	0	0	–	0	0	0	1.5
KNN	1	1	0.5	0	1	–	0.5	0	4
XGBoost	1	1	0.5	0	1	0.5	–	0	4
DNN	1	1	1	1	1	1	1	–	7

Table 3.19: Performance comparison of different classifiers for DS4 using the accumulated preference index.

Classifier	**LR**	**DT**	**RF**	**SVM**	**NB**	**KNN**	**XGBoost**	**DNN**	AP_i
LR	–	1	1	1	1	1	0.5	0	5.5
DT	0	–	0	0	0	0	0	0	0
RF	0	1	–	0.5	1	1	0	0	3.5
SVM	0	1	0.5	–	1	1	0	0	3.5
NB	0	1	0	0	–	0.5	0	0	1.5
KNN	0	1	0	0	0.5	–	0	0	1.5
XGBoost	0.5	1	1	1	1	1	–	0	5.5
DNN	1	1	1	1	1	1	1	–	7

The accumulated preference index for all classifiers across four datasets (DS1, DS2, DS3, and DS4) is illustrated in Figure 3.6. The deep neural network consistently achieved the highest AP_i values, demonstrating superior performance and robustness in all datasets. Its dominance was particularly evident in DS3 and DS4, where it excelled in accuracy, precision, F1-score, and sensitivity, showcasing its versatility across various evaluation criteria. Meanwhile, XGBoost also demonstrated strong performance, ranking second in DS1 and DS4, while LR and SVM showed moderate reliability on simpler datasets. RF performed consistently but did not surpass the DNN or XGBoost in any scenario. NB presented a trade-off, achieving the highest sensitivity in DS2 but the lowest precision; this makes it more suitable for sensitivity-focused applications.

As a result, the DNN emerged as the most reliable and effective classifier, particularly for balanced datasets like DS4. While the comprehensive performance of the DNN makes it a suitable choice, models such as XGBoost and SVM may be more appropriate for tasks that

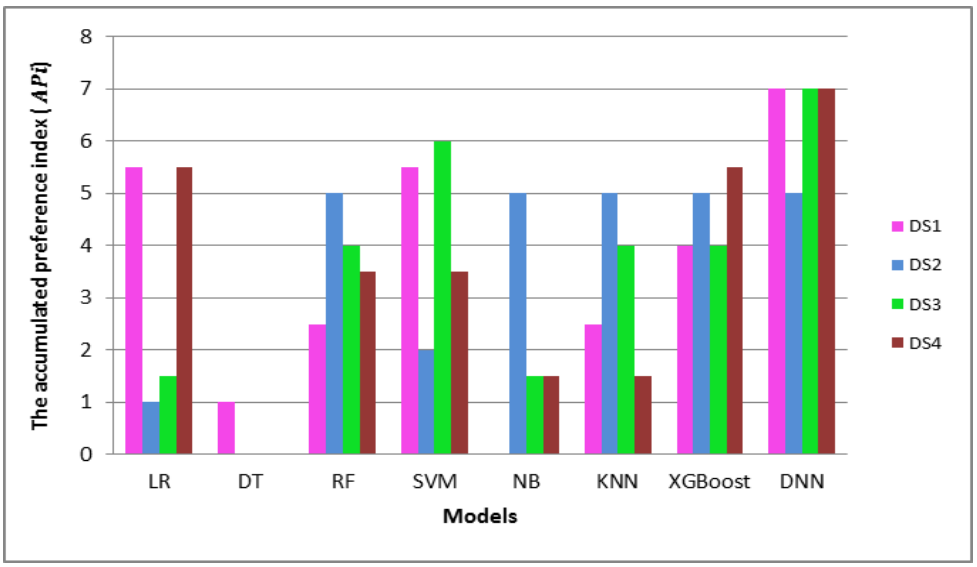


Figure 3.6: Performance comparison of the top five models for DS1 dataset.

prioritize precision or specificity. Figure 3.6 emphasizes the dominance of the DNN and the complementary strengths of other models in specific contexts.

3.4 Comparison With the State-of-the-Art Methods

This section presents a comparative analysis of our proposed approach with state-of-the-art studies published between 2018 and 2024, using identical protocols, datasets, and performance metrics. As shown in Table 3.20, our method consistently surpassed all previous research efforts. Our method achieved peak accuracies of 99.30% on DS1, 98.97% on DS2, 89.47% on DS3, and 95.83% on DS4, outperforming existing SOTA methods across all four datasets. These findings underscore the effectiveness and robustness of our approach in addressing the challenges associated with breast cancer classification.

Table 3.20: Performance comparison of the proposed method with SOTA methods.

Refs	Year	Dataset	Method	Accuracy (%)
[25]	2021	DS1(WDBC)	Hierarchical Clustering Random Forest (HCRF)	97.05
[29]	2020		SVM	97.66
[44]	2021		KNN	98.00
[41]	2018		quadratic kernel-based SVM	98.10
[55]	2019		Linear SVM	99.00
[31]	2023		MLJAR, Lazy Predict	99.12
[43]	2020		FCLF-CNN	99.28
Ours	**2024**		**DNN with (WFS+SMOTE+HPO)**	**99.30**
[39]	2020	DS2 (WBCD)	RBFNN	97.00
[29]	2020		Random Forest	97.01
[56]	2018		Naïve Bayes	97.36
[257]	2018		LR	98.10
[31]	2023		Lazy Predict	98.54
[258]	2022		(SVM+LR+NB+DT) + ANN	98.83
Ours	**2024**		**DNN with (WFS + RUS + HPO)**	**98.97**
[30]	2021	DS3 (WPBC)	FW-KNCM + BOA + RDF	80.00
[58]	2022		Multi-stage learning technique	82.00
[31]	2023		Orange	83.50
Ours	**2024**		**DNN with (WFS+RUS+HPO)**	**89.47**
[31]	2023	DS4 (Breast cancer Coimbra dataset)	MATLAB Classification Learner	91.3
Ours	**2024**		**DNN with (WOFS+HPO)**	**95.83**

3.5 Conclusion

This chapter highlights the potential benefits of using a multitasking framework for breast cancer classification. The first task involved a feature selection process to extract the most relevant features, which enhanced the quality of representation across the four datasets used. The second task focused on balancing the data to address the class imbalance, using both oversampling and undersampling techniques. Specifically, the SMOTE and KNNOR methods were employed to generate synthetic instances of the minority class, while the random undersampling technique was used to reduce the size of the majority class to achieve balance. This phase significantly enhanced the balance of the datasets, leading to improved model performance. The third task involved hyperparameters optimization to fine-tune each model's parameters, ensuring they reached optimal values for maximum performance. In total, seven widely used machine learning algorithms, along with deep neural networks models, were tested and evaluated using various performance metrics. Integrating these processes within the multitasking framework produced exceptional results, surpassing state-of-the-art techniques across all datasets. These findings highlight the effectiveness and robustness of the proposed approach in tackling the challenges of breast cancer classification.

Chapter 4

BREAST CANCER CLASSIFICATION ENHANCEMENT USING DEEP-MODIFIED TRANSFER LEARNING

4.1 Introduction

This chapter aims to develop a precise and automated method for breast cancer classification, leveraging deep modified transfer learning with pretrained CNNs, including ResNet50, MobileNetV2, DenseNet121, and Xception, to effectively extract relevant features from the breast ultrasound images dataset. The extracted features are passed to high-performance classifiers, including SVM, KNN, XGBoost, and Softmax, to classify the extracted features as benign or malignant accurately. The results demonstrate the proposed method's superiority by benchmarking its performance against existing methods using identical datasets, experimental protocols, and performance metrics, contributing to significant improvements in early breast cancer classification and facilitating the development of more personalized treatment strategies.

4.2 Proposed System

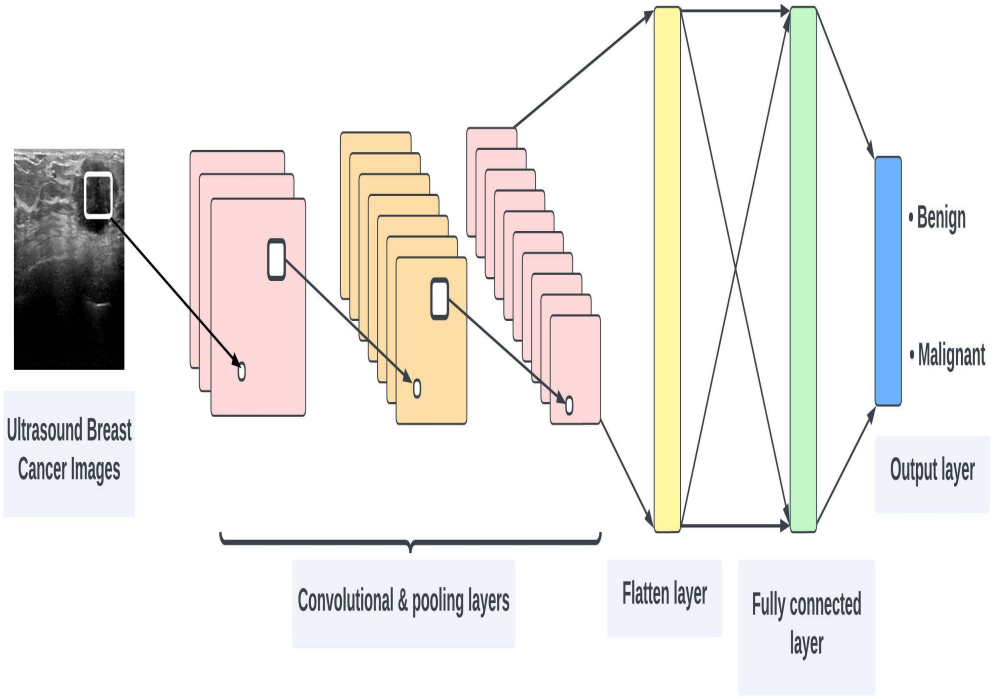
The proposed methodology for breast cancer classification from ultrasound images utilizes two distinct approaches: (a) a custom-built CNNs approach and (b) a TL-based CNN approach, as illustrated in Figure 4.1. The custom-built approach (Figure 4.1a) involves designing and de-

veloping three CNN architectures from scratch to process the ultrasound breast images. These models comprise multiple convolutional and pooling layers for feature extraction, followed by a flattening layer that prepares these features for subsequent fully connected layers. The network culminates in an output layer that classifies the input images as benign or malignant. This approach emphasizes learning features directly from the dataset, focusing on the specific characteristics present in breast ultrasound images.

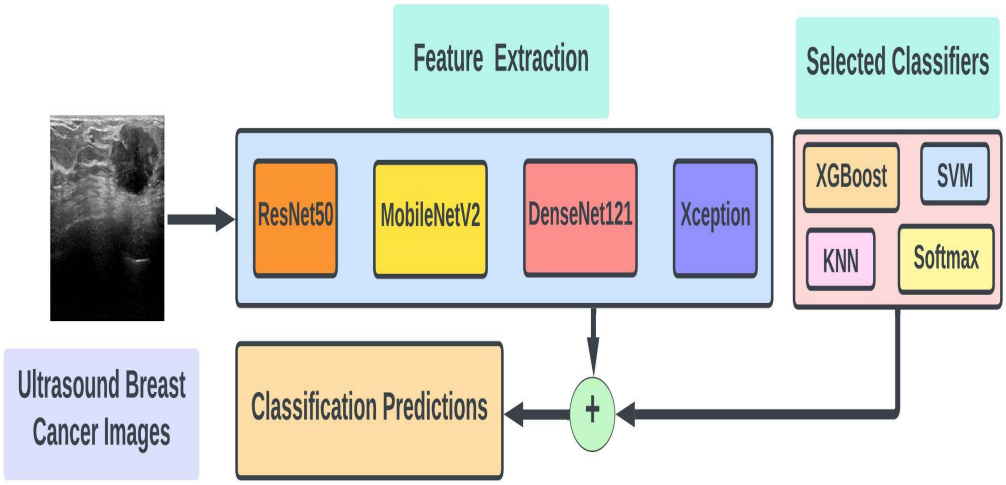
The second approach (Figure 4.1b) uses transfer learning to enhance classification accuracy. It leverages several well-established pre-trained CNN architectures, including ResNet50, MobileNetV2, DenseNet121, and Xception. These models capture the most relevant features of breast ultrasound images, enabling efficient feature extraction. The extracted features are then input into advanced machine learning classifiers—SVM, KNN, XGBoost, and Softmax. Each classifier is meticulously trained and tested to determine the optimal configuration for accurately distinguishing between benign and malignant tumors. Extensive experiments and validations were analyzed to assess the performance of the custom-built CNN models and the TL-based frameworks. Metrics such as accuracy, precision, and F1-score are employed to demonstrate their robustness and generalizability. The combination of both approaches provides a comprehensive solution for breast cancer classification, offering a reliable tool for medical practitioners and aiding in accurate classification and treatment planning.

4.2.1 Data Description

In this study, the BUSI dataset [113] was employed to evaluate the efficacy of the proposed methodology. This challenging dataset comprises 780 ultrasound images of women 25-75 years of age, each resized to 500 × 500 pixels in PNG format. The dataset was categorized into three distinct classes: benign (487 samples), malignant (210 samples), and normal (177 samples). The images were acquired using advanced ultrasound scanning systems—LOGIQ E9 and LOGIQ E9 Agile—at Baheya Hospital in Cairo, Egypt, adhering to standardized imaging protocols to maintain consistency and quality. The dataset was partitioned into two subsets: 80% for training and 20% for testing. This split ensured sufficient data for the models to learn while retaining an adequate portion for assessing their effectiveness. Figure 4.2 presents representative samples from each class in the BUSI dataset, highlighting the variability and complexity of the images used in this study.

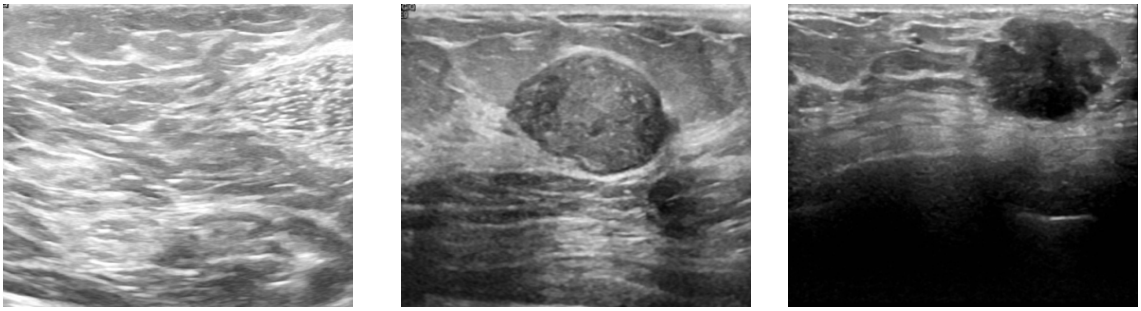


(a) Scratched-built CNNs approach.



(b) Transfer learning approach.

Figure 4.1: Flowchart diagrams and the proposed breast cancer classification method.



(a) Normal (b) Benign (c) Malignant

Figure 4.2: Representative samples of breast ultrasound images: normal, benign, and malignant classifications.

4.2.2 Costumized CNN Architectures

Several convolutional neural network architectures were developed from scratch with their structures, as illustrated in Figure 4.3. We conducted numerous experiments by adjusting the number of convolution and max-pooling layers and the number and size of filters in each layer. All custom-built CNNs share the same fundamental layer types, including convolutional layers, max-pooling, flattened, and dense layers. For clarity, we structured the CNN architectures into blocks consisting of convolutional and max-pooling layers. The input dimensions of the custom-built CNNs are 224×224 pixels. After processing through the blocks in each model, the output sizes for configurations (a), (b), and (c) are reduced to 53×53, 23×23, and 8×8 pixels, respectively. The convolutional layers employed the Rectified Linear Unit (ReLU) activation function, which was chosen for its computational efficiency and ability to mitigate vanishing gradient issues. We selected an Adam optimizer with a cross-entropy loss function to ensure stable and efficient convergence during training. This comprehensive experimentation with architectural variations provided valuable insights into the design principles that drive the performance improvement of custom-built CNNs.

4.2.3 Transfer Learning-Based CNN Models

Transfer learning is a technique that leverages pre-trained models, which were originally developed for one task, to manage another related task. This approach reduces training time and computational complexity because the model benefits from prior knowledge gained from a similar problem. In image classification, numerous CNN architectures are designed and optimized through benchmark datasets like ImageNet which have demonstrated exceptional performance and adaptability when employed via transfer learning. In this study, we used four state-of-the-

art transfer learning models, namely, ResNet50, MobileNetV2, DenseNet121, and Xception to extract the most significant features from the BUSI dataset. This study identified and used the most important features for classifying breast cancer by taking advantage of the strengths of different architectures.

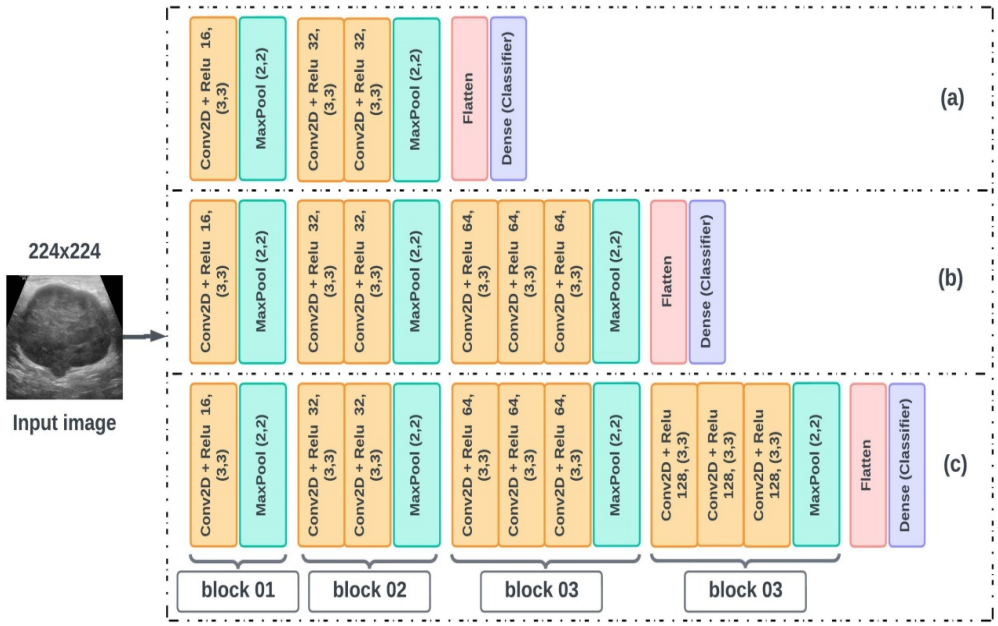


Figure 4.3: Architectural designs of custom-built CNNs for breast ultrasound image classification.

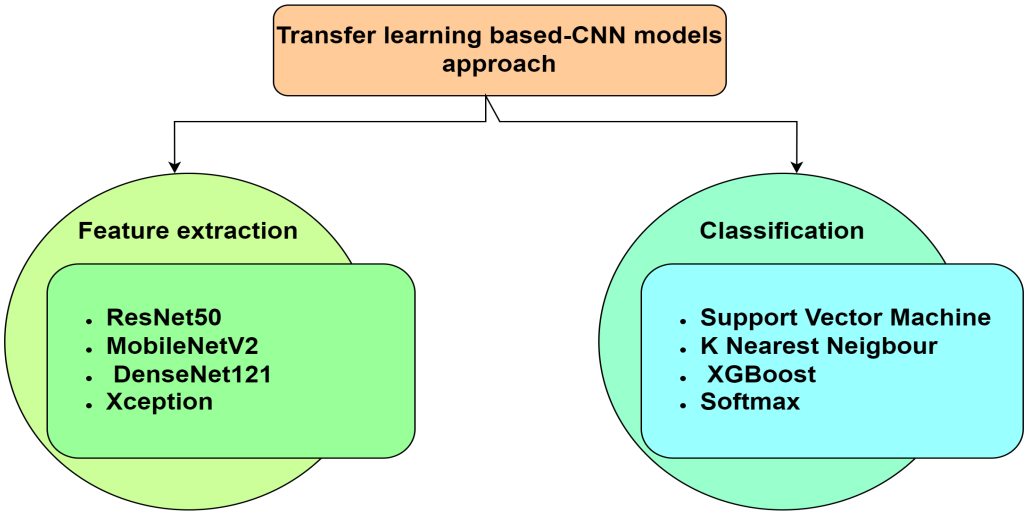


Figure 4.4: Deep feature extraction with machine learning classifiers for breast cancer classification.

4.2.3.1 ResNet50

ResNet50 is a deep convolutional neural network architecture consisting of 50 layers, first introduced by He et al. [259]. It comprises 48 convolutional layers, one max pooling layer, and one average pooling layer. As one of the most widely used versions of ResNet, it accepts an input size of 256×256 pixels. In 2015, ResNet achieved remarkable success in the ImageNet challenge, outperforming other architectures, and has been employed in various computer vision tasks, including object detection, image segmentation, and image classification [260]. The incorporation of skip connections between layers, as illustrated in Figure 4.5, not only reduces the overall number of parameters but mitigates the vanishing gradient problem commonly encountered in very deep CNNs [120, 153, 261].

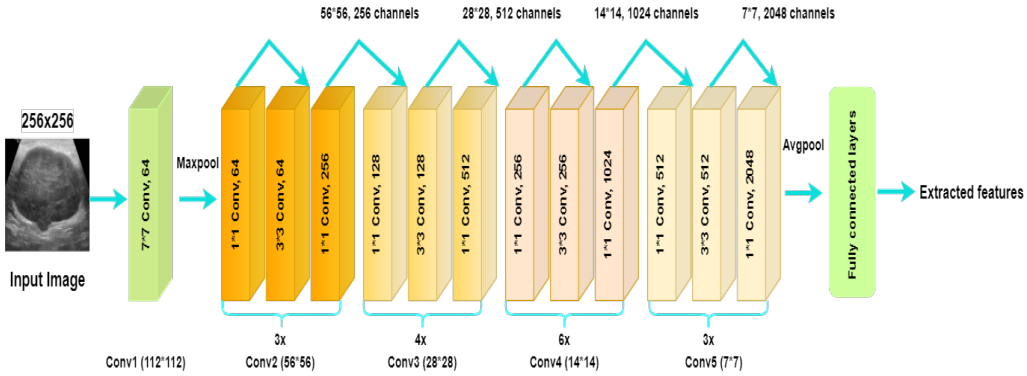


Figure 4.5: Architectural design of ResNet50 for efficient feature extraction in breast cancer classification.

4.2.3.2 MobileNetV2

MobileNetV2 is a convolutional neural network architecture developed by Google specifically for mobile and embedded vision applications [262, 263]. This architecture improves upon MobileNetV1, emphasizing enhanced efficiency and speed compared to traditional computer vision models. MobileNetV2 employs depthwise separable convolutions, which enhance computational efficiency over standard convolutions, thereby minimizing the number of parameters and Floating-Point Operations (FLOPS) required for inference. In addition, MobileNetV2 incorporates a linear bottleneck layer at the end of each residual block, as illustrated in Figure 4.6, which further reduces the computational demands during inference [264].

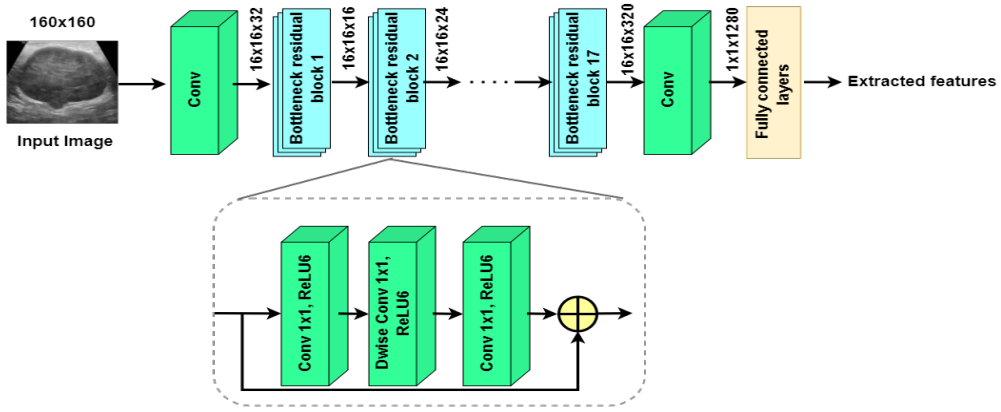


Figure 4.6: Architectural design of MobileNetV2 for efficient feature extraction in breast cancer classification.

4.2.3.3 DenseNet121

DenseNet121 is a convolutional neural network architecture developed by Gao Huang and colleagues in 2017 [265]. It aims to enhance traditional models such as ResNet and Inception by introducing a novel layer type known as a "dense block." This dense block facilitates more efficient information flow between various layers within the network, effectively addressing the vanishing gradient problem and significantly reducing the number of learnable parameters at each layer [120, 153]. As a result, DenseNet121 has demonstrated superior accuracy and faster training times compared to other architectures, as illustrated in Figure 4.7.

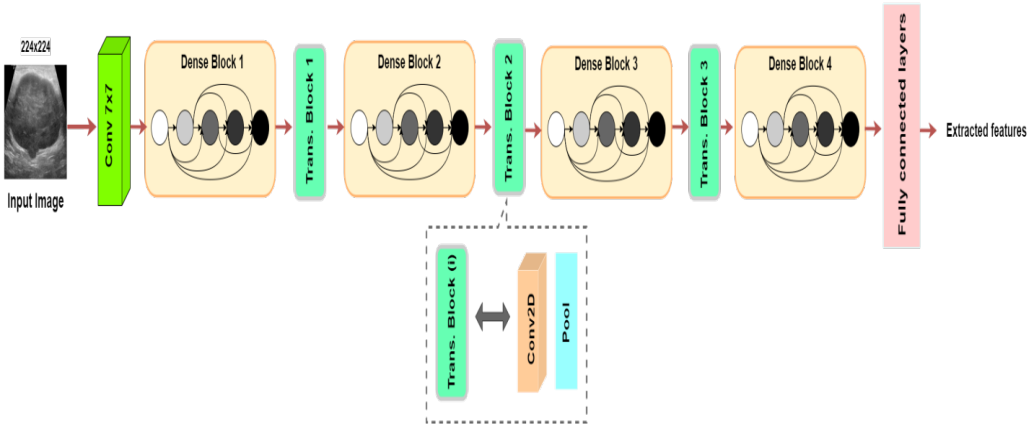


Figure 4.7: Architectural design of DenseNet121 for efficient feature extraction in breast cancer classification.

4.2.3.4 Xception

The Xception model, also known as Extreme Inception, was initially introduced by F. Chollet [266] to enhance the Inception-V3 architecture by utilizing depthwise separable convolutions. This approach allows for the independent processing of spatial and channel dimensions of images during training [153, 267]. The model has found significant application and success in various fields of medical imaging. Figure 4.8 illustrates the architecture of the Xception feature extractor.

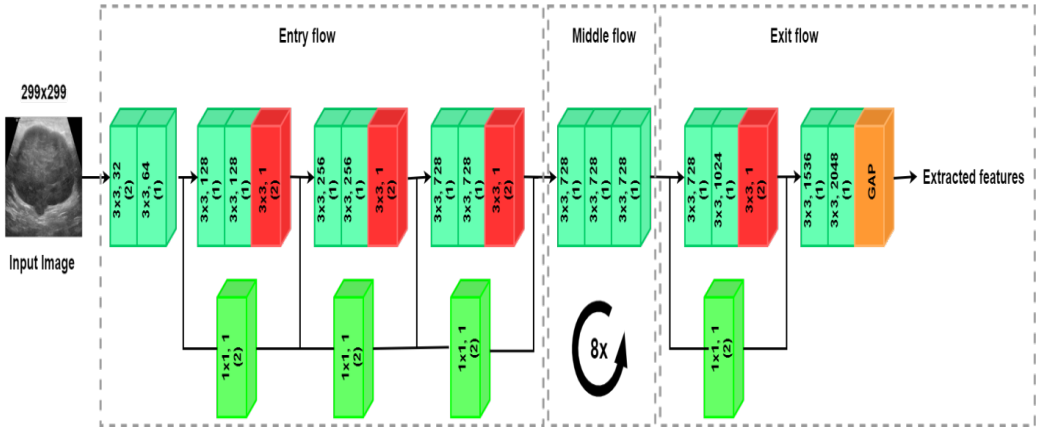


Figure 4.8: Architectural design of Xception for efficient feature extraction in breast cancer classification.

4.2.4 Machine Learning Classifiers

The features extracted from various pre-trained models, as described in the previous section, were assessed using several widely used classifiers: SVM, KNN, XGBoost, and Softmax, to achieve effective classification of breast cancer tumors. Numerous experiments have been performed using various classifiers across different transfer learning models to determine the best-performing classifier for each corresponding pre-trained model.

The support vector machines, introduced by V. Vapnik et al. in 1995 [268], is a supervised machine learning technique used for classification tasks by constructing a hyperplane to distinguish between two or more classes. The k-nearest neighbors [269] is an effective machine learning classifier for various classification tasks. KNN does not require a training phase; instead, it relies on a distance function to assess inputs based on the similarity of data instances and their features.

Moreover, we employed the extreme gradient boosting algorithm, developed by Chen and Guestrin in 2016 [270], as a high-performance classifier. XGBoost integrates multiple weak

decision trees with low-performance rates to produce a robust model. The last layer of each pre-trained model, which serves as the classifier, was modified by adjusting the number of neurons to align with the data targets. In this case, Softmax was used with two units corresponding to the number of labels in the dataset.

4.3 Result and Discussion

To accurately evaluate the performance of the proposed method, a series of experiments have been conducted to analyze the effectiveness of various pre-trained CNNs paired with different classifiers. This approach aims to identify the optimal classifier and pre-trained model based on their performance. To achieve this, several widely recognized evaluation criteria have been applied, including accuracy, precision, and F1-score.

4.3.1 Classification With Costumized CNN Models

The performance of three custom-built CNN models: Model_01, Model_02, and Model_03, was assessed based on accuracy, precision, and F1-score, as outlined in Table 4.1. Each model underwent testing under various configurations, varying the number of training epochs (15 and 100) and batch sizes (16, 32, and 64). Among the three models, Model_03 showed exceptional performance, achieving peak values of 90.70% in accuracy, precision, and F1-score when trained for 100 epochs with a batch size of 64. Model_02 also showed strong overall results, with an accuracy of 89.92%, precision of 89.87%, and an F1-score of 89.89%, all attained under the same 100-epoch training regimen but with a batch size of 32. In contrast, Model_01 consistently delivered more moderate performance across all configurations tested, reaching a maximum accuracy of 85.27%. These findings highlight the significance of hyperparameter optimization in enhancing CNN performance.

4.3.1.1 Impact of Model Complexity on Performance

The relationship between model complexity and performance is highlighted in Table 4.2, which presents the number of trainable parameters and the corresponding testing accuracy for three custom-built CNN architectures. As the complexity of the models increases, measured by the number of trainable parameters, there is a clear improvement in testing accuracy. The Model_01, with 104,225 trainable parameters, achieves an accuracy of 85.27%. The Model_02, with 140,545 parameters, performs better, attaining an accuracy of 89.92%. Notably, the

Model_03, the most complex with 483,905 trainable parameters, delivers the highest testing accuracy of 90.70%. These findings demonstrate that increasing model complexity, when carefully managed, can significantly enhance performance.

Table 4.1: Performance evaluation of three scratch-built CNN models using different performance metrics.

Model	Epochs	Batch Size	Accuracy	Precision	F1-score
Model_01		16	0.8527	0.8630	0.8553
	15	32	0.8527	0.8504	0.8501
		64	0.8217	0.8200	0.8136
	100	16	0.8527	0.8593	0.8547
Model_02		16	0.8605	0.8591	0.8596
	15	32	0.8682	0.8745	0.8700
		64	0.8527	0.8508	0.8512
	100	32	0.8992	0.8987	0.8989
Model_03		16	0.7829	0.7844	0.7635
	15	32	0.7674	0.8271	0.7199
		64	0.7984	0.7967	0.7858
	100	**64**	**0.9070**	**0.9070**	**0.9070**

Table 4.2: The number of trainable parameters vs. testing accuracy of scratch-built CNN models.

Model	Number of Trainable Parameters	Accuracy (%)
Model_01	104,225	85.27
Model_02	140,545	89.92
Model_03	**483,905**	**90.70**

4.3.2 Performance Assessment of Deep Modified Transfer Learning-Based CNN Networks

This sub-section examines the performance of various deep-modified transfer learning-based CNNs, specifically ResNet50, MobileNetV2, DenseNet121, and Xception, utilizing a range of classifiers, including SVM, KNN, XGBoost, and Softmax. As outlined in Table 4.3, each classifier exhibited unique characteristics, resulting in disparate levels of performance enhancement

across the models evaluated. The Softmax classifier consistently surpassed the other classifiers using accuracy and precision metrics.

The ResNet50 demonstrated robust performance across all classifiers; however, applying the Softmax classifier markedly enhanced its metrics, culminating in an accuracy of 87.59%, a precision of 84.21%, and an F1-score of 80.00%. These results represent a substantial improvement compared to the alternative classifiers, which achieved lower accuracies of 82.95% with SVM, 77.52% with KNN, and 79.84% with XGBoost. Furthermore, MobileNetV2 exhibited exceptional performance when integrated with the Softmax classifier, achieving an accuracy of 93.79%. This performance significantly outstripped that of the SVM (78.29%), KNN (75.19%), and XGBoost (76.74%). Notably, both accuracy and F1-score metrics for MobileNetV2 experienced significant enhancements with the Softmax classifier, reaching 94.73% and 90.00%, respectively, thereby underscoring the classifier's efficacy in amplifying the model's output. The Xception model attained an accuracy of 93.02%, with precision and F1-score metrics of 94.59% and 88.60%, respectively, when utilizing the Softmax classifier.

Although Xception demonstrated acceptable performance with SVM and XGBoost, achieving accuracies of 84.50% and 82.95% respectively, it encountered difficulties with KNN, recording only 70.54%. This observation reiterates the crucial role of the Softmax classifier in realizing the full potential of the Xception model. Lastly, DenseNet121 produced robust results across all classifiers, achieving particularly impressive outcomes with the Softmax classifier. It recorded the highest accuracy of 95.34% among all models and classifiers assessed, yielding accuracy and F1-score values of 90.90% and 93.02%, respectively. This finding highlights the vital role of the Softmax classifier in enhancing DenseNet121's precision in our breast cancer tumor classification task.

In conclusion, the Softmax classifier has proven to be instrumental in enhancing the performance of deep-modified transfer learning-based CNN architectures, with DenseNet121 achieving a noteworthy accuracy of 95.34%. This analysis emphasizes the critical importance of selecting the appropriate classifier to optimize CNN models for breast cancer classification, with Softmax emerging as the superior choice within this evaluation.

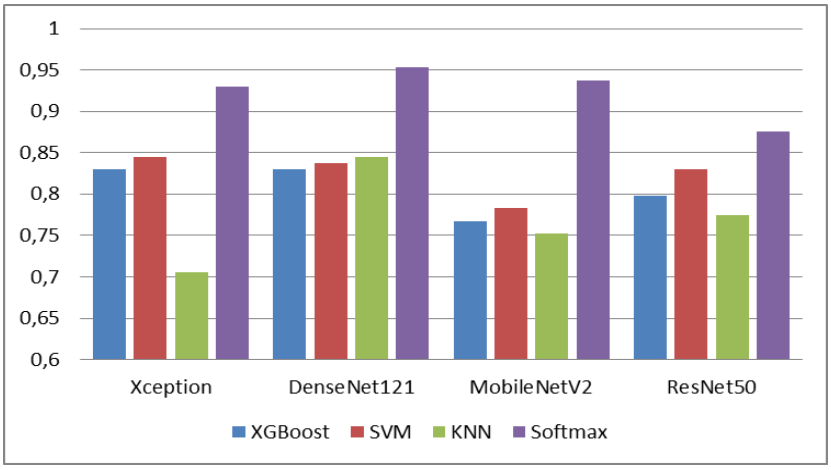
The performance of four classifiers—Softmax, KNN, SVM, and XGBoost—has been evaluated across four deep learning models: Xception, DenseNet121, MobileNetV2, and ResNet50, utilizing three key metrics: accuracy, precision, and F1-score, as illustrated with 3D bar charts in Figure 4.9. Notably, Softmax consistently outperforms the other classifiers across all metrics, achieving high rates in accuracy and F1-score, highlighting its robustness in classification tasks. SVM also shows competitive performance, especially in terms of Precision. This visualization

effectively emphasizes the strengths of each classifier-model combination, offering valuable insights for enhancing classification performance with advanced deep learning architectures.

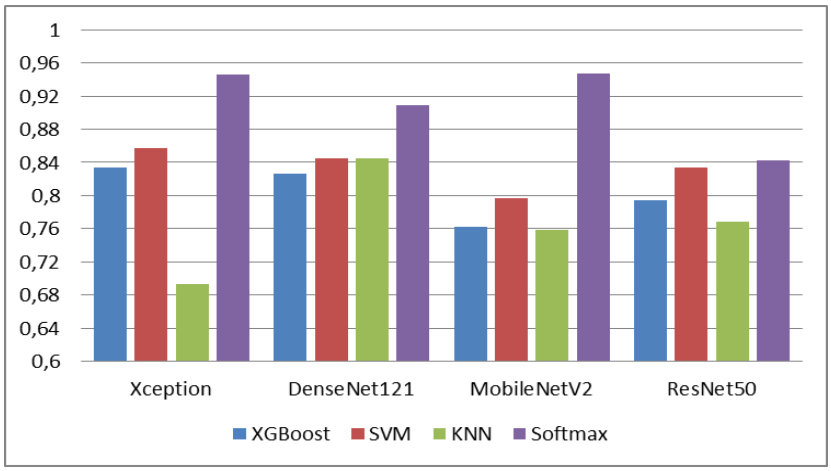
Table 4.3: Performance comparison of deep CNN networks using SVM, KNN, XGBoost, and Softmax classifiers.

Deep CNN models	**Machine learning classifiers**	**Accuracy (%)**	**Precision (%)**	**F1-score (%)**
ResNet50	SVM	82.95	83.33	81.88
	KNN	77.52	76.85	76.93
	XGBoost	79.84	79.46	78.82
	Softmax	87.59	84.21	80.00
MobileNetV2	SVM	78.29	79.65	75.66
	KNN	75.19	75.88	75.47
	XGBoost	76.74	76.17	74.99
	Softmax	93.79	94.73	90.00
DenseNet121	SVM	83.72	84.48	82.60
	KNN	84.50	84.48	83.87
	XGBoost	82.95	82.65	82.71
	Softmax	**95.34**	**90.90**	**93.02**
Xception	SVM	84.50	85.71	83.32
	KNN	70.54	69.31	69.64
	XGBoost	82.95	83.33	81.88
	Softmax	93.02	94.59	88.60

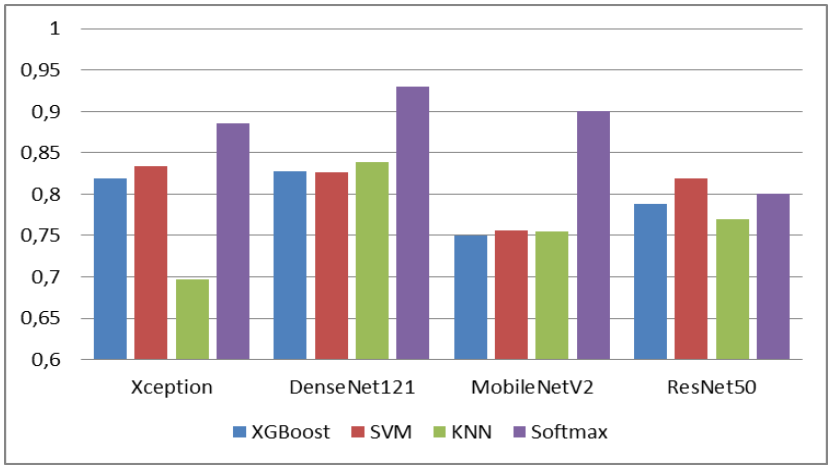
Figure 4.10 presents a performance comparison of four CNN models—ResNet50, MobileNetV2, DenseNet121, and Xception—when combined with the Softmax classifier. The evaluation utilized accuracy, precision, and F1-score as metrics. DenseNet121 outperformed the other models across all metrics, showcasing its remarkable ability to extract and classify critical features while maintaining an impressive balance between accuracy and precision. MobileNetV2 also demonstrated strong results, particularly in Precision, coming close to DenseNet121, making it an efficient option for minimizing false positives in medical applications. ResNet50 achieved high accuracy but fell short in precision and F1-score, indicating potential challenges with false predictions. Xception displayed competitive accuracy but had a relatively lower F1-score. This analysis underscores DenseNet121's superiority and highlights the trade-offs associated with the other models in breast cancer classification tasks.



(a) Accuracy



(b) Precision



(c) F1-score

Figure 4.9: Performance comparison of different feature extractors and classifiers based on several evaluation criteria.

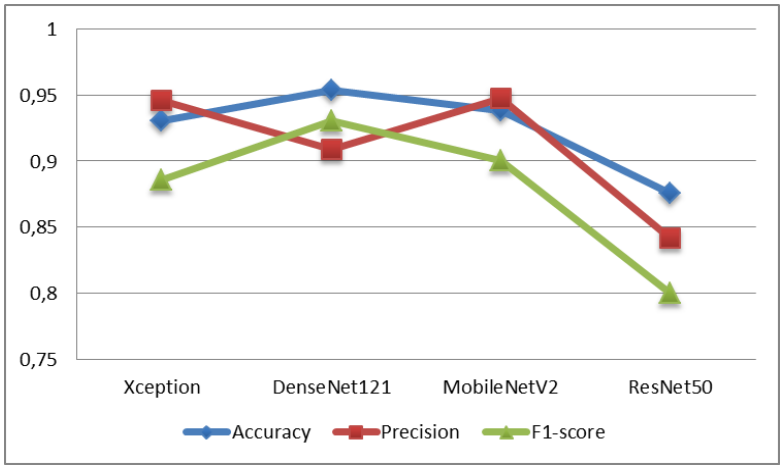


Figure 4.10: Performance comparison of different feature extractors with the softmax classifier.

4.3.3 Performance Comparison

In this subsection, we compared the performance of four transfer learning models—ResNet50, MobileNetV2, DenseNet121, and Xception—against the scratch-built CNN model (Model_03) using accuracy, precision, and f1-score as evaluation metrics (see Figure 4.11). DenseNet121, combined with the Softmax classifier, achieves superior performance across all metrics, surpassing both the pre-trained models and the scratch-built CNN model. This highlights the effectiveness of transfer learning in leveraging pre-trained knowledge for feature extraction and classification, especially when compared to models built from scratch, which typically require extensive datasets and training to achieve comparable results. The analysis underscores the importance of pre-trained models in scenarios with limited data availability.

4.4 Comparison With the State-of-the-Art Methods

In this section, we compare our proposed method to the state-of-the-art approaches that utilize the same dataset, protocol, and performance metrics. As shown in Table 4.4, our Deep Modified DenseNet121 Network achieves the highest accuracy of 95.34%, surpassing all previous studies. The closest competing methods include the DeepCls approach by X. Qi et al. [271], which achieved an accuracy of 94.51%. These studies, conducted between 2020 and 2023, illustrate the effectiveness of various deep learning techniques applied to the BUSI dataset. Our approach surpasses these methods by utilizing a refined architecture specifically designed to enhance performance in breast ultrasound image classification. This significant improvement highlights the robustness and efficacy of our model in achieving superior classification accuracy.

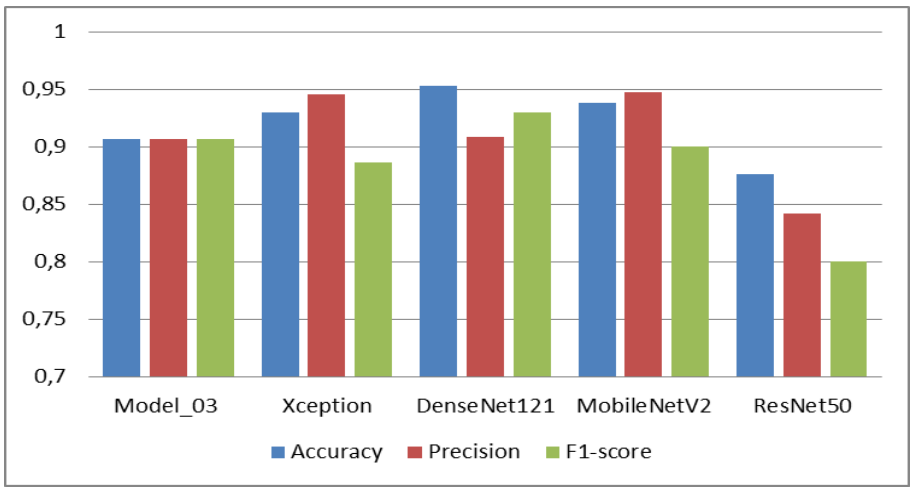


Figure 4.11: Performance comparison of different approaches based on different performance criteria.

Table 4.4: Performance comparison of the proposed method with SOTA methods.

Refs	Year	Dataset	Method	Accuracy (%)
[272]	2021	BUSI	TL-based deep representation scaling (DRS)	91.5
[120]	2022		ResNet50	73.72
[271]	2022		DeepCls	94.51
Ours	**2024**	**BUSI**	**Deep Modified DenseNet121 Network**	**95.34**

4.5 Conclusion

This study proposes a method using deep-modified transfer learning-based CNN networks for enhanced breast cancer classification and treatment of breast cancer a feasible method. First, the BUSI images are inputted into feature extractors to capture their relevant features. Subsequently, the extracted features are accurately classified as benign or malignant using high-performance classifiers such as SVM, KNN, XGBoost, and softmax. Using deep modified Densenet121 with softmax classifier outperforms other CNN networks and existing SOTA techniques, achieving 95.34% (accuracy), 90.90% (precision), and 93.02% (F1-score). The research outcomes offer substantial societal and academic benefits by facilitating breast cancer classification and advancing research in medical imaging and diagnostics. However, the study has certain limitations, including the use of a specific dataset and the necessity for real-world testing to ensure practical viability. Future plans include integrating the method with other diagnostic modalities and exploring applications beyond breast cancer to enhance diagnostic capabilities.

Chapter 5

DEEP CNN FEATURE FUSION AND LASSO-BASED FEATURE SELECTION FOR BREAST CANCER CLASSIFICATION

5.1 Introduction

This chapter presents a comprehensive approach for accurate breast tumor classification using ultrasound images by integrating multiple transfer learning models. First, multiple pre-trained CNN architectures, including MobileNetV2, DenseNet121, and InceptionV3, are utilized to extract complementary features from the ultrasound images, capturing a broader range of relevant information compared to single-model approaches. The features extracted by each model are then combined to create a more robust feature representation, mitigating potential overfitting. Then, we apply the LASSO regression method to select the most significant features from the combined features vector, further enhancing model efficiency and generalization. This integrated approach achieves a high classification performance, outperforming previously reported results and demonstrating its effectiveness in improving diagnostic accuracy. The resulting methodology forms the basis of an efficient decision support system designed to assist clinicians in making more informed diagnostic decisions regarding breast tumors.

5.2 Proposed System

A comprehensive schematic of the proposed system for breast cancer classification is depicted in Figure 5.1, highlighting a robust architecture aimed at delivering highly accurate BC classification. This system effectively utilizes transfer learning models for extract and integrate crucial features from breast cancer images, enhancing predictive performance. First, breast cancer images are inputted into several pre-trained CNNs, including MobileNetV2, DenseNet121, and InceptionV3. These models, recognized for their exceptional feature extraction abilities, analyze the input images to capture essential features, which are flattened into one-dimensional vectors. The extracted feature vectors from each CNN are merged into a single feature vector, consolidating the diverse and overlapping information obtained from each model. This fusion ensures a comprehensive representation that encompasses a rich array of features, providing an ultimate understanding of the breast cancer images. The integrated feature vector is subsequently passed through a series of fully connected layers, functioning as dense neural networks. These layers are designed to identify intricate patterns and relationships within the combined features, effectively reducing dimensionality while augmenting the model's ability to differentiate between classes. Finally, the output from the fully connected layers is directed into a dense classification layer, which acts as the decision-making component, predicting whether the breast cancer image is benign or malignant.

The proposed system has undergone extensive validation using a variety of datasets, hyperparameter configurations, and experimental setups. This rigid evaluation demonstrates its high accuracy and strong generalization capability, inducting the system as a reliable and effective tool for breast cancer classification. By leveraging advanced transfer learning techniques and a thoughtfully designed architecture, this method significantly advances breast cancer classification.

5.2.1 Dataset Description

The dataset used to evaluate the proposed approach is described in Section 4.2.1.

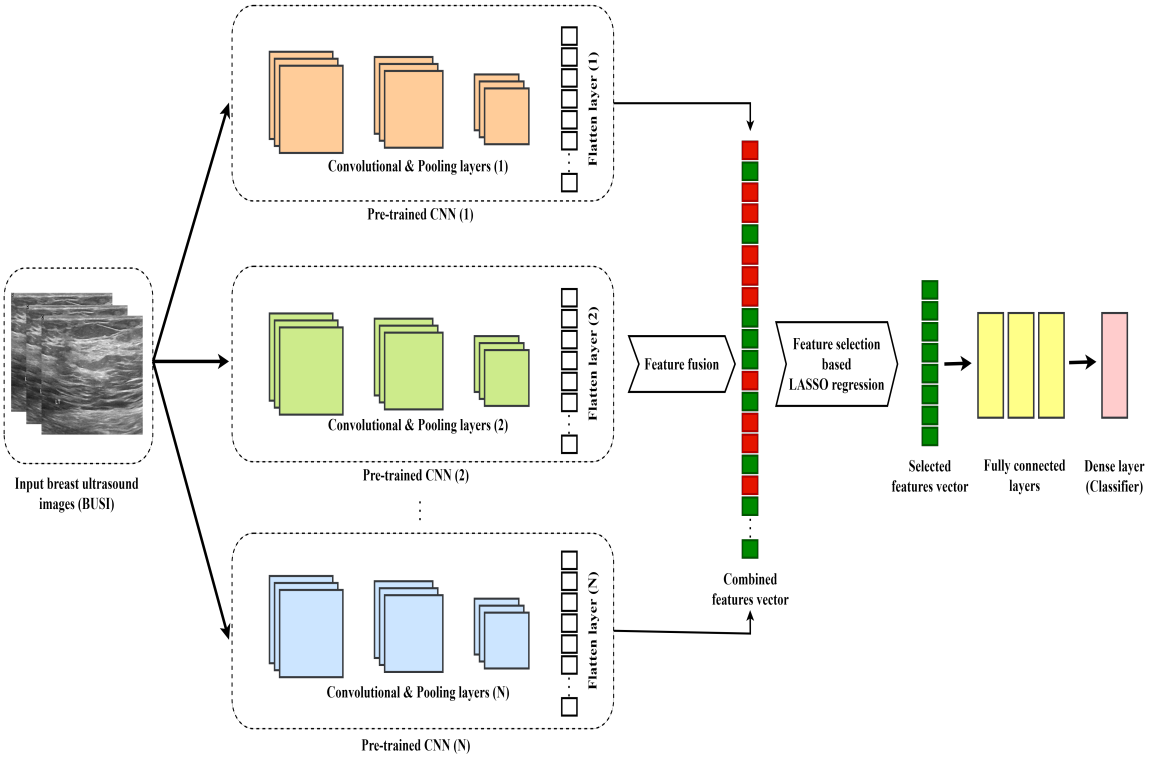


Figure 5.1: Schematic diagram of the proposed approach.

5.2.2 Transfer Learning

Transfer learning is a machine learning strategy that leverages knowledge gained from a source task to improve performance on a related target task [273]. This approach is particularly beneficial when labeled data for the target task is limited. By utilizing knowledge learned from a source task with abundant labeled data, TL mitigates the risk of overfitting and enhances the model's ability to generalize to the target task [274]. Furthermore, TL is highly effective in achieving strong performance even with small datasets, as it allows models to leverage pre-learned features, reducing the need for extensive training [275]. In this study, three deep CNN architectures were implemented using transfer learning: MobileNetV2 (Section 4.2.3.2), DenseNet121 (Section 4.2.3.3), and InceptionV3.

5.2.2.1 InceptionV3

InceptionV3 is a deep convolutional neural network comprising 48 layers, known for its versatility and effectiveness in various research applications (see Fig. 5.2) [276]. As part of the Inception family, InceptionV3 has been pre-trained on the ImageNet dataset, allowing it to achieve outstanding accuracy across multiple tasks. This model is particularly noted for its

computational efficiency, balancing resource usage and performance, making it a desired choice for applications such as image classification [277].

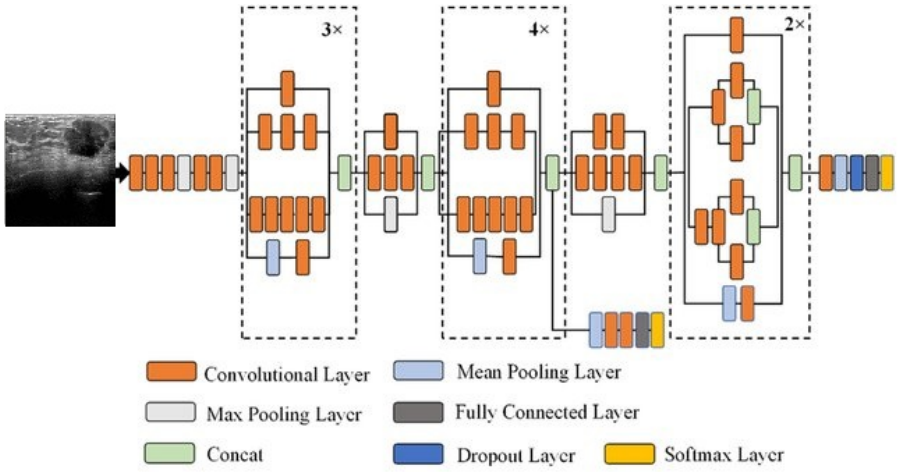


Figure 5.2: Architectural design of InceptionV3 for efficient feature extraction in breast cancer classification.

5.2.3 Concatenating Pretrained CNNs

Combining features from multiple pre-trained CNNs is an effective strategy to enhance performance across various applications [278]. This technique involves extracting multi-model features, consisting of representations derived from the individual features of each CNN model [279]. Integrating diverse features captured from each CNN model can provide a richer and more comprehensive representation of the input data, often leading to improved classification accuracy in the target task [280, 281].

5.2.4 LASSO-Based Feature Selection

The LASSO is a regression technique introduced by Tibshirani in 1996 [282] that effectively combines parameter estimation with feature selection. LASSO employs a form of penalized regression that applies an L1 penalty to the regression coefficients. This penalty drives some coefficients to shrink to zero, thereby eliminating irrelevant or redundant features and ensuring optimal feature selection. Mathematically, LASSO minimizes the residual sum of squares while setting a constraint on the sum of the absolute values of the coefficients, as the following [283]:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \sum_{i=1}^{N} \left(y_i - \alpha - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\}, \quad \text{s.t.} \sum_{j=1}^{p} |\beta_j| \leq t \tag{5.1}$$

Where y represents the dependent variable, x denotes the independent variables, α is the intercept term, and β are the coefficients for the p predictors. The constant t serves as the penalty parameter, where smaller values of t encourage greater sparsity by setting more coefficients to zero. When $t < t_0$, irrelevant features are effectively excluded from the model [284].

LASSO improves both classification accuracy and model interpretability by effectively combining the advantages of ridge regression and subset selection. Its capability to manage multicollinearity makes it particularly beneficial for datasets with highly correlated predictors, as it selects a single representative feature while shrinking others to zero. It is valuable, particularly in high-dimensional datasets where both interpretability and the reduction of multicollinearity are crucial. LASSO has proven invaluable in identifying key features across clinical, imaging, and genomic data. Its unique ability to eliminate irrelevant variables while maintaining predictive power has solidified LASSO's role as a fundamental tool in feature selection for contemporary machine learning pipelines. Tables 5.1 and 5.2 illustrate the effect of LASSO regression-based feature selection on diminishing the dimensionality of feature vectors for various CNN models.

Table 5.1: The effect of FS-based LASSO method on the number of selected features for individual models.

Pre-trained CNNs	**N. of Features**	**Round 1**			**Round 2**		
		Alpha	**Selected Features**	**Accuracy (%)**	**Alpha**	**Selected Features**	**Accuracy (%)**
MobileNetV2	62,721	**0.001**	**631**	**96.15**	**0.02**	**264**	**96.15**
		0.01	**427**	**96.15**	0.03	149	91.54
		0.1	10	83.85	0.04	81	89.23
					0.05	59	87.69
					0.002	**593**	**96.15**
					0.003	564	94.62
					0.004	543	95.38
					0.005	**515**	**96.15**
DenseNet121	50,177	0.001	616	93.08	0.02	269	90.00
		0.01	410	93.08	0.03	179	92.31
		0.1	21	82.31	0.04	115	86.15
					0.05	82	87.69
				95.38	**0.002**	**584**	**95.38**
					0.003	551	94.62
					0.004	529	93.85
					0.005	507	91.54
InceptionV3	51,201	0.001	598	93.85	0.002	559	93.85
		0.01	367	90.00	0.003	527	93.85
		0.1	10	84.62	**0.004**	**492**	**94.62**
					0.005	**465**	**94.62**

Table 5.2: The effect of FS-based LASSO method on the number of selected features for different pre-trained CNNs combinations.

Pre-trained CNNs	**N. of Features**	**Round 1**			**Round 2**		
		Alpha	**Selected Features**	**Accuracy (%)**	**Alpha**	**Selected Features**	**Accuracy (%)**
MobileNetV2 + DenseNet121	112,896	0.001	626	95.38	0.02	285	94.62
		0.01	**421**	**96.92**	0.03	180	92.31
		0.1	18	86.92	0.04	109	88.46
					0.05	82	85.38
MobileNetV2 + InceptionV3	113,920	0.001	642	89.23	**0.02**	**258**	**93.08**
		0.01	410	91.54	0.03	164	89.23
		0.1	15	84.62	0.04	111	87.69
					0.05	65	92.31
DenseNet121 + InceptionV3	101,376	0.001	628	89.23	0.02	273	91.54
		0.01	405	93.08	**0.03**	**188**	**94.62**
		0.1	16	86.92	0.04	119	90.00
					0.05	78	87.69
MobileNetV2 + DenseNet121 + InceptionV3	164,096	0.001	645	96.15	**0.02**	**280**	**96.92**
		0.01	**431**	**96.92**	0.03	195	91.54
		0.1	18	89.23	0.04	116	87.69
					0.05	85	86.92

5.3 Results and Discussion

To evaluate the effectiveness of transfer learning, we analyzed the performance of three pre-trained CNN models: MobileNetV2, DenseNet121, and InceptionV3, using common metrics such as accuracy, sensitivity, specificity, precision, and F1-score. as presented in Table 5.3. Moreover, the features extracted from these models were fused, enabling the combination of complementary information, addressing overfitting, and enhancing the feature space. The combined feature vector is refined using the LASSO method, which selects the most significant features, reducing redundancy and improving computational efficiency. In this study, the BUSI dataset was divided into training and testing subsets with an 80%-20% split, ensuring a robust evaluation of the proposed method.

5.3.1 Performance Assessment of Pretrained CNNs

We used three pre-trained CNNs, including MobileNetV2, DenseNet121, and InceptionV3, which were chosen for their architectural diversity and proven track record in image classification tasks, particularly in medical imaging. Each model was trained and tested separately under various configurations, including different hyperparameters such as epochs, batch size, and learning rate, to ensure a comprehensive comparison of performance.

Among the models evaluated, MobileNetV2 showed superior performance across most metrics. It achieved the highest accuracy at 95.34%, surpassing both DenseNet121 and InceptionV3. This indicates that MobileNetV2 maintains high performance, making it an excellent choice for breast cancer classification tasks. It demonstrated remarkable precision and specificity, both at 100%, showcasing its effectiveness in accurately identifying true positives while minimizing false positives. Additionally, it attained an F1-score of 90.90%, highlighting its well-balanced performance in managing both precision and sensitivity.

The DenseNet121 model demonstrated admirable performance with an accuracy of 93.79%, slightly following behind MobileNetV2. Its dense connectivity architecture enables efficient gradient propagation throughout the network, helping it capture complex patterns within the dataset. However, its sensitivity was somewhat lower at 83.33%, indicating that it may have missed some positive cases. On the positive side, DenseNet121 achieved a high specificity of 98.85%, highlighting its effectiveness in minimizing false positives.

Ultimately, InceptionV3 achieved an impressive accuracy of 90.69%, demonstrating its effectiveness in capturing multi-scale features for this task. However, compared to MobileNetV2, InceptionV3 showed slightly lower precision and sensitivity, particularly in cases where the learning rate and batch size were not optimal. This underscores the importance of careful hyperparameter tuning to extract the best performance of each model. Table 5.3 presents a comprehensive overview of accuracy, precision, sensitivity, specificity, and F1-score for each pre-trained CNN, highlighting their strengths and weaknesses in breast cancer classification.

Table 5.3: Performance assessment of different pre-trained CNN models.

CNN Models	Epochs	Batch size	Learning rate	Performance (%)				
				Accuracy	Precision	Sensitivity	Specificity	F1-score
MobileNetV2	20	16	0.01	92.24	97.05	78.57	98.85	86.84
			0.001	93.02	97.14	80.95	98.85	88.31
			0.0001	93.79	97.22	83.33	98.85	89.74
			0.00001	89.92	93.93	73.80	97.70	82.66
		32	0.01	93.79	97.22	83.33	98.85	89.74
			0.001	94.57	**100.0**	83.33	**100.0**	90.90
			0.0001	92.24	97.05	78.57	98.85	86.84
			0.00001	87.59	86.11	73.80	94.25	79.48
		64	0.01	92.24	92.10	83.33	96.55	87.50
			0.001	93.02	**100.0**	78.57	**100.0**	88.00
			0.0001	92.24	94.44	80.95	97.70	87.17
			0.00001	82.17	78.78	61.90	91.95	69.33
	100	**32**	**0.001**	**95.34**	97.36	**88.09**	98.85	**92.50**
DenseNet121	20	16	0.01	88.37	93.54	69.04	97.70	79.45
			0.001	89.92	89.18	78.57	95.40	83.54
			0.0001	90.69	96.87	73.80	**98.85**	83.78
			0.00001	88.37	90.90	71.42	96.55	80.00
		32	0.01	90.69	87.50	83.33	94.25	85.36
			0.001	93.79	90.47	90.47	95.40	90.47
			0.0001	89.92	93.93	73.80	97.70	82.66
			0.00001	86.04	92.85	61.90	97.70	74.28
		64	0.01	87.59	88.23	71.42	95.40	78.94
			0.001	89.14	96.66	69.04	**98.85**	80.55
			0.0001	89.92	93.93	73.80	97.70	82.66
			0.00001	82.94	88.46	54.76	96.55	67.64
	100	**32**	**0.001**	**94.57**	**97.29**	**85.71**	**98.85**	**91.13**
InceptionV3	20	16	0.01	89.14	81.81	**85.71**	90.80	83.72
			0.001	88.37	84.61	78.57	93.10	81.48
			0.0001	**90.69**	89.47	80.95	95.40	**85.00**
			0.00001	87.59	90.62	69.04	96.55	78.37
		32	0.01	89.92	85.36	83.33	93.10	84.33
			0.001	**90.69**	**94.11**	76.19	97.70	84.21
			0.0001	89.14	88.88	76.19	95.40	82.05
			0.00001	82.94	85.71	57.14	95.40	68.57
		64	0.01	89.14	86.84	78.57	94.25	82.50
			0.001	87.59	88.23	71.42	95.40	78.94
			0.0001	89.14	91.17	73.80	96.55	81.57
			0.00001	75.19	77.77	33.33	95.40	46.66
	100	**32**	**0.001**	**90.69**	**94.11**	76.19	**97.70**	84.21

5.3.2 Enhanced Performance Through Feature Fusion of Pretrained CNN Models

This section examines the potential of feature fusion by combining the features from three high-performing pre-trained CNN models: MobileNetV2, DenseNet121, and InceptionV3. By integrating features from these models, we aim to leverage their strengths and improve overall performance. The results demonstrate that this approach can significantly enhance performance. Among the tested combinations, the fusion of MobileNetV2 and DenseNet121 achieved

the best results with an impressive accuracy of 96.12% along with high scores across other evaluation metrics, outperforming both individual models and other combinations as summarized in Table 5.4. These findings underline the effectiveness of feature fusion as a strategy to boost classification performance by capturing complementary information from different models.

Table 5.4: Performance of concatenated pre-trained CNN models.

Pre-trained CNNs	**Performance (%)**				
	Accuracy	**Precision**	**Sensitivity**	**Specificity**	**F1-score**
MobileNetV2	95.34	97.36	**88.09**	98.85	92.50
DenseNet121	94.57	97.29	85.71	98.85	91.13
InceptionV3	90.69	94.11	76.19	97.70	84.21
MobileNetV2+ DenseNet121	**96.12**	**100.0**	**88.09**	**100.0**	93.67
MobileNetV2+ InceptionV3	94.57	94.87	**88.09**	97.70	91.35
DenseNet121+ InceptionV3	**96.12**	97.43	90.47	98.85	**93.82**
MobileNetV2+ DenseNet121+ InceptionV3	95.34	97.36	**88.09**	98.85	92.5

5.3.3 Comparative Analysis of LASSO-Based Feature Selection for Pretrained CNN Combinations

The experimental results of applying LASSO-based feature selection to various combinations of pre-trained CNN models are summarized in Table 5.5. As shown in the table, LASSO-based FS consistently enhances the performance of most model combinations by refining the feature space and reducing redundancy. Notably, the combinations of MobileNetV2 and DenseNet121, and MobileNetV2+ InceptionV3 achieved the highest performance with all metrics, reaching 99.23% accuracy, 97.29% precision, 100% sensitivity, 98.93% specificity, and a 98.63% F1-score, indicating exceptional classification performance. The combination of all three models—MobileNetV2, DenseNet121, and InceptionV3—also yielded similarly impressive results, achieving the same metrics values as the MobileNetV2 + DenseNet121 combination. However, the combination of DenseNet121 and InceptionV3 exhibited slightly reduced performance with 96.92% accuracy, 90.00% precision, 100% sensitivity, 95.74% specificity, and 94.73% F1-score. Figures 5.6-5.9 clearly demonstrate the effectiveness of LASSO-based FS in improving the predictive power of fused CNN features, while also highlighting that the specific choice of models for combination can significantly impact the final performance.

Table 5.5: Performance of pretrained CNN combinations with LASSO-based FS method.

Pre-trained CNNs	**Performance (%)**				
	Accuracy	**Precision**	**Sensitivity**	**Specificity**	**F1-score**
MobileNetV2	**99.23**	**97.29**	**100.0**	**98.93**	**98.63**
DenseNet121	98.46	97.22	97.22	98.93	97.22
InceptionV3	96.15	87.80	100.0	94.68	93.50
MobileNetV2+ DenseNet121	**99.23**	**97.29**	**100.0**	**98.93**	**98.63**
MobileNetV2+ InceptionV3	**99.23**	**97.29**	**100.0**	**98.93**	**98.63**
DenseNet121+ InceptionV3	96.92	90.00	**100.0**	95.74	94.73
MobileNetV2+ DenseNet121+ InceptionV3	**99.23**	**97.29**	**100.0**	**98.93**	**98.63**

The performance of various models and their combinations is analyzed before and after LASSO-based feature selection using key metrics: accuracy, sensitivity, specificity, precision, and F1-score, as illustrated in Figures 5.3-5.9. FS consistently enhances performance, particularly in accuracy, sensitivity, and F1-Score, with significant gains noted in models like MobileNetV2. The combined models, especially MobileNetV2 + DenseNet121 and MobileNetV2 + InceptionV3 exhibit balanced improvements across all metrics, highlighting the benefits of model combinations. This analysis emphasizes the importance of FS in optimizing model performance in complex scenarios.

Figure 5.3: Impact of LASSO-based feature selection on the performance of the MobileNetV2 model in breast cancer classification.

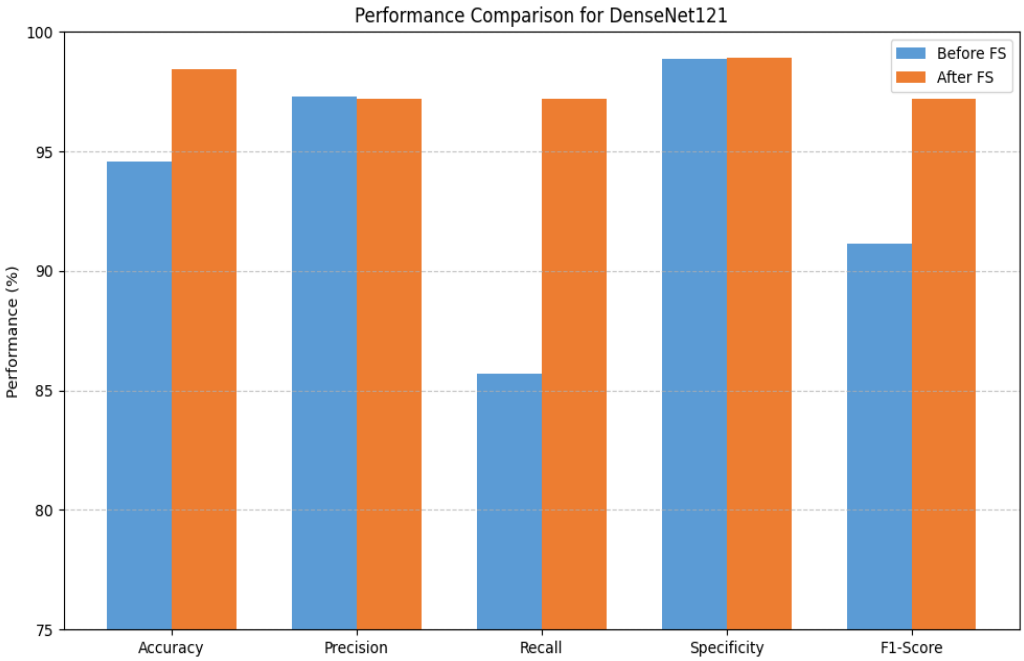


Figure 5.4: Impact of LASSO-based feature selection on the performance of the DenseNet121 model in breast cancer classification.

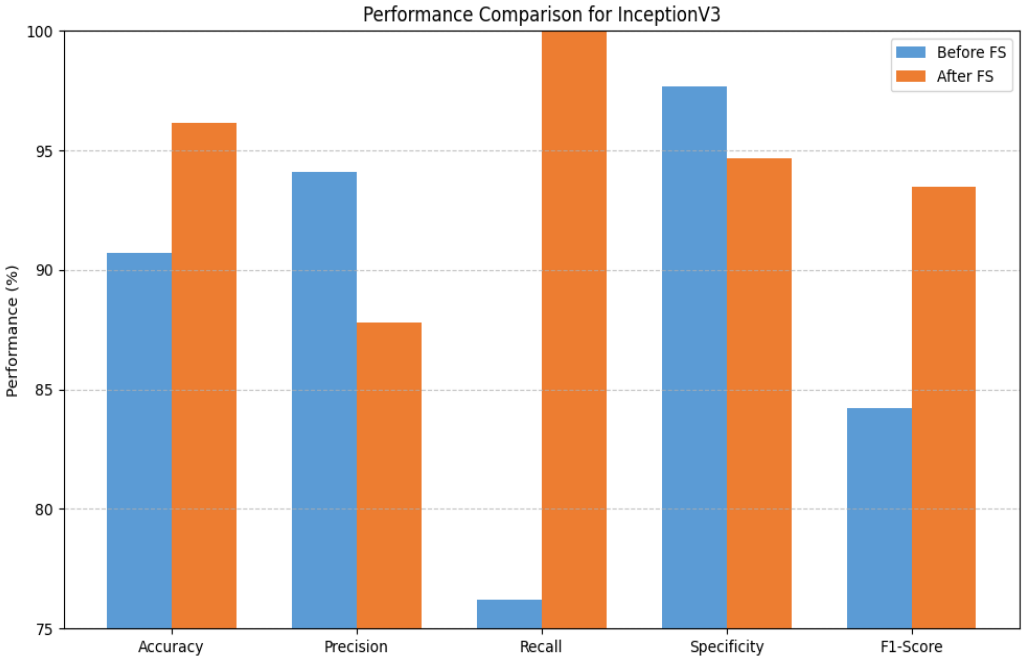


Figure 5.5: Impact of LASSO-based feature selection on the performance of the InceptionV3 model in breast cancer classification.

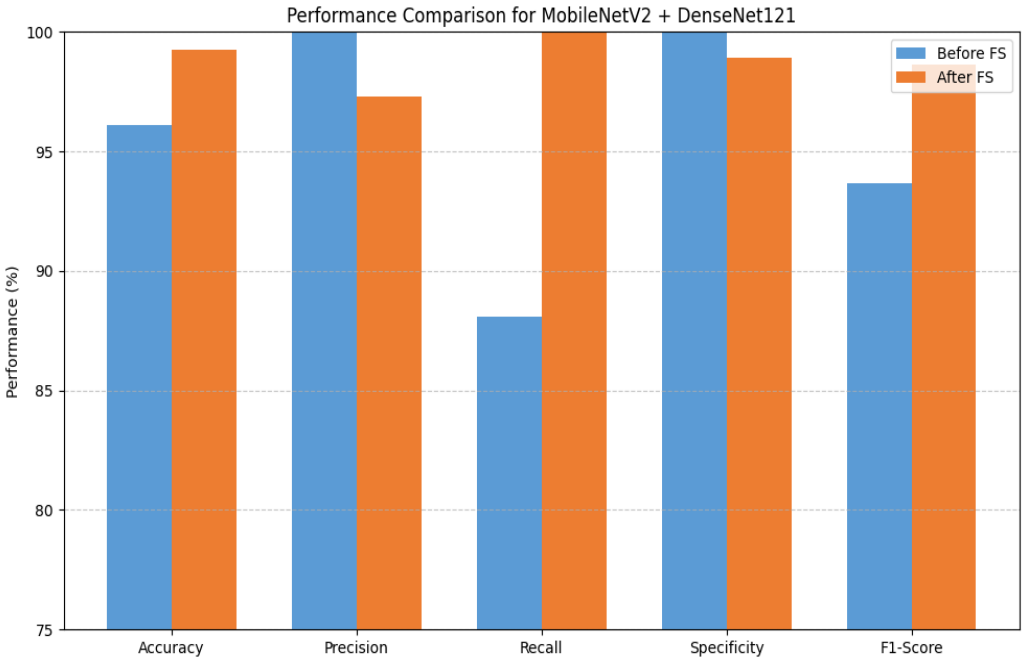


Figure 5.6: Impact of LASSO-based feature selection on the performance of the MobileNetV2 + DenseNet121 combination in breast cancer classification.

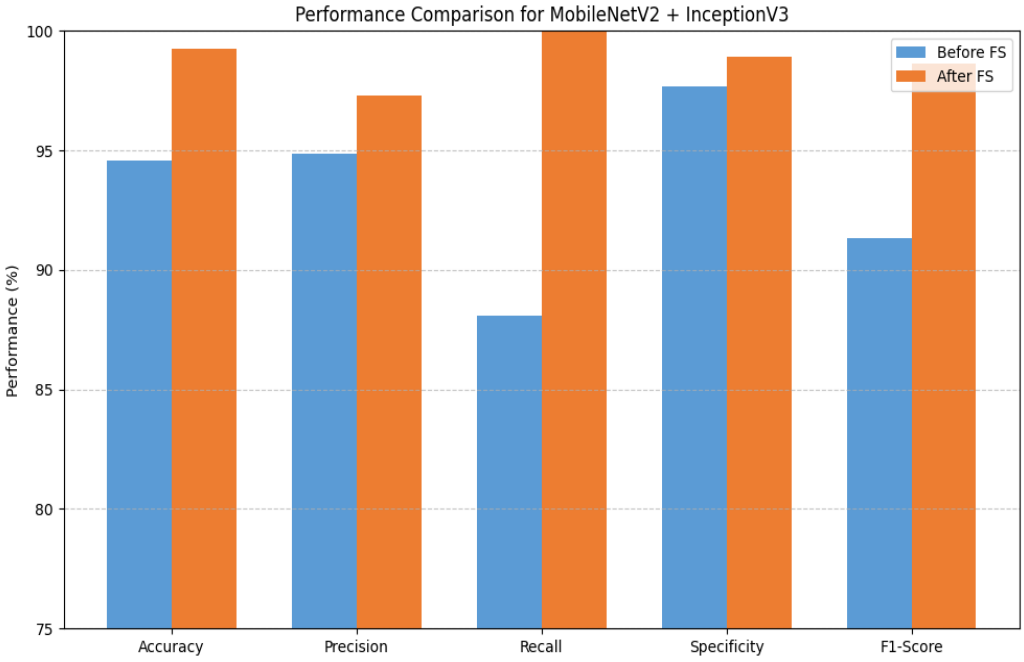


Figure 5.7: Impact of LASSO-based feature selection on the performance of the MobileNetV2 + InceptionV3 combination in breast cancer classification.

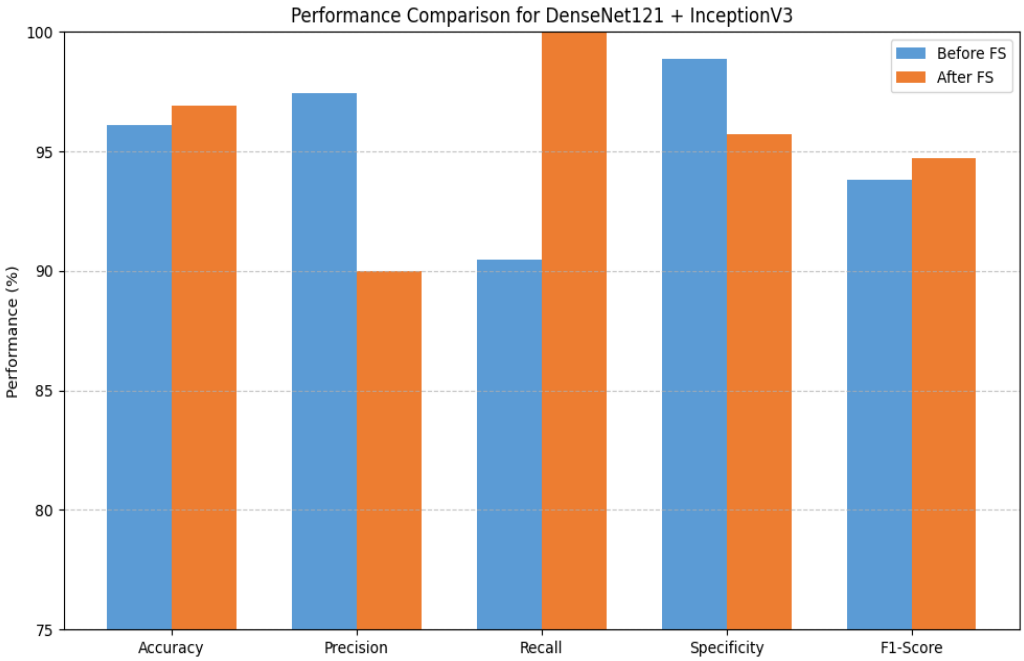


Figure 5.8: Impact of LASSO-based feature selection on the performance of the DenseNet121 + InceptionV3 combination in breast cancer classification.

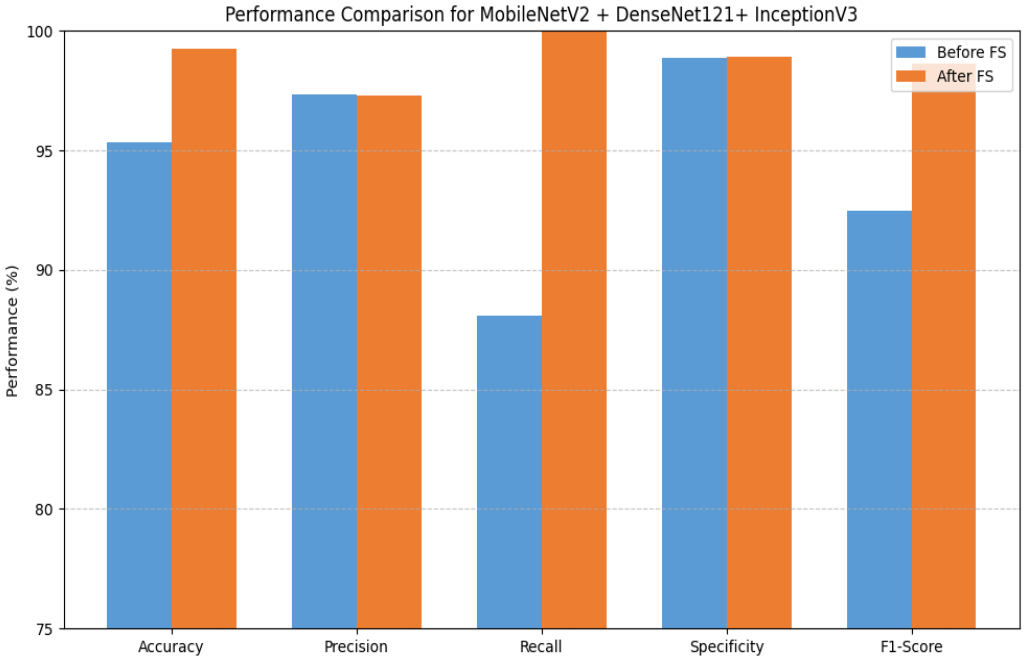


Figure 5.9: Impact of LASSO-based feature selection on the performance of the MobileNetV2 + DenseNet121 + InceptionV3 combination in breast cancer classification.

The proposed method for breast cancer classification employs LASSO-based feature selection in the combinations of MobileNetV2 and DenseNet121, MobileNetV2+ InceptionV3, and MobileNetV2+ DenseNet121+ InceptionV3 models, showcasing outstanding performance across essential evaluation metrics. As illustrated in Figure 5.10, the confusion matrix reflects impressive specificity and sensitivity, with only one false positive and no false negatives, leading to reliable classification results. The ROC curve further demonstrates the effectiveness of the proposed model, achieving an impressive area under the curve (AUC) of 0.99 (see Figure 5.11).

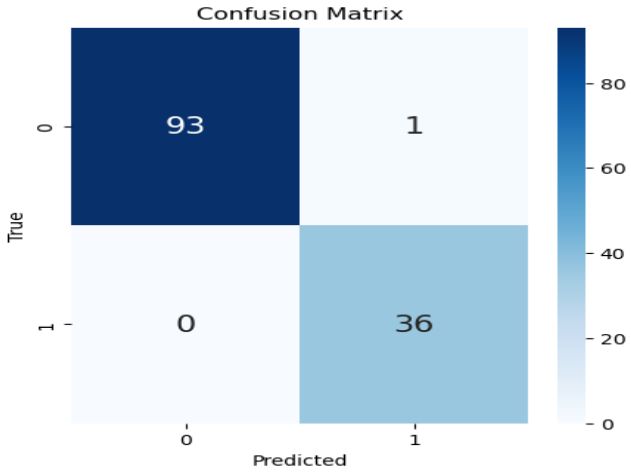


Figure 5.10: Confusion matrix of the best models using LASSO-based feature selection for breast cancer classification.

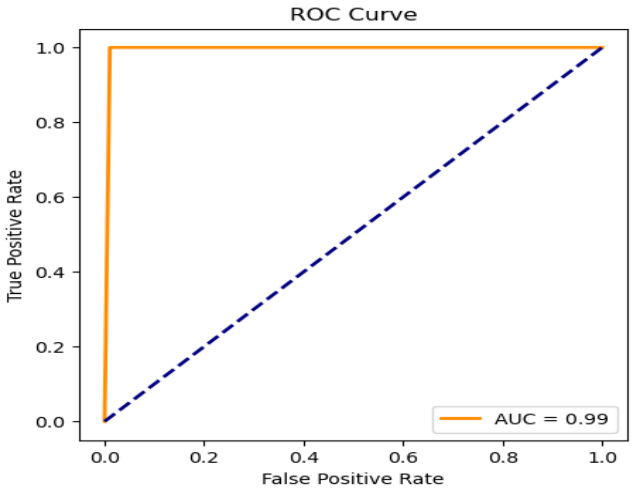


Figure 5.11: ROC curve of the best models using LASSO-based feature selection for breast cancer classification.

5.4 Comparison With the State-of-the-Art Methods

In this section, we compare our proposed method with several recent studies that have utilized similar protocols and assessment metrics, as shown in Table 5.6. Our method achieved remarkable performance, attaining an impressive accuracy of 99.23%, which surpasses all competing studies. The closest competitors, Qi et al. [271] and Wang et al. [285], recorded accuracies of 94.51% and 94.52%, respectively. Despite utilizing advanced deep learning techniques, their performance did not match the capabilities of our model. Our proposed method, which integrates MobileNetV2 and DenseNet121 with LASSO-based feature selection, shows substantial enhancements. These findings underscore the robustness and effectiveness of our approach in breast cancer classification, establishing a new benchmark for accuracy in this domain.

Table 5.6: Performance comparison of the proposed method with SOTA methods.

Refs	Year	Dataset	Method	Accuracy (%)
[272]	2021	BUSI	Transfer learning	91.5
[271]	2022	BUSI	DeepCls	94.51
[285]	2021	BreaKHis	Deep FE-BkCapsNet	94.52
[147]	2021	BreaKHis	PTGAN	90
Ours	**2024**	**BUSI**	**MobileNetV2+DenseNet121 with FS based LASSO regression**	**99.23**

5.5 Conclusion

This chapter introduces an innovative approach to accurately classifying breast cancer images. Rather than relying on a single pre-trained CNN model, which may not capture sufficient information for the head classifier, we utilize three highly effective pre-trained models: MobileNetV2, DenseNet121, and InceptionV3. These models collaboratively extract significant features from breast cancer images, creating a comprehensive, high-dimensional feature vector that leverages the strengths of each model while mitigating overfitting. Subsequently, we employ LASSO-based feature selection to identify and retain only the most relevant features from this combined vector. Our proposed method has demonstrated superior performance compared to existing SOTA techniques, achieving an impressive accuracy of 99.23%.

Chapter 6

CONCLUSION

6.1 Contributions Summary

This thesis has investigated integrating advanced machine learning and deep learning techniques to address critical limitations in traditional breast cancer diagnostic methods. By leveraging state-of-the-art datasets and computational models, this research significantly enhances the accuracy and efficiency of breast cancer classification. The key contributions are summarized as follows:

- Demonstrated the transformative potential of deep learning, particularly convolutional neural networks, in analyzing complex medical imaging data, showcasing their ability to extract intricate features relevant to breast cancer classification.
- Highlighted the effectiveness of transfer learning in enhancing classification accuracy, particularly in scenarios with limited datasets, demonstrating its ability to leverage pre-trained knowledge for improved performance.
- Emphasized the importance of feature selection, data balancing, and hyperparameter optimization techniques for developing robust and generalizable models, ensuring reliable performance across diverse patient populations, and reducing bias.
- Developed CAD systems capable of reducing radiologists' workload while providing reliable preliminary assessments, streamlining the diagnostic workflow, and potentially improving diagnostic speed.
- Affirmed improved performance rates and reduced false positives and negatives, leading to more timely and accurate clinical decisions and improving patient outcomes.

6.2 Difficulties and Working Environment

While this research successfully achieved its objectives, several challenges were encountered. Working with medical imaging data presented inherent difficulties, primarily concerning data availability due to patient privacy regulations and data-sharing agreements. Variations in image quality across different imaging devices and institutions necessitated meticulous preprocessing to ensure data consistency and model robustness. The demand for high computational resources also posed a significant limitation, particularly when training complex deep learning models, which require substantial processing power and memory capacity. Furthermore, ensuring the generalizability of the developed models across diverse patient populations was a critical challenge, necessitating careful selection and rigorous validation of representative datasets to mitigate bias and enhance clinical applicability.

6.3 Perspectives and Future Plans

This thesis contributes to the ongoing efforts in breast cancer research by developing CAD classification systems that harness the power of ML and DL. As these technologies continue to improve, they promise to revolutionize breast cancer diagnostics, resulting in better patient management and increased survival rates. Future work should focus on several key areas:

- Enhancing the developed models through architectural improvements, advanced training techniques, and exploring innovative deep learning architectures.
- Exploring the integration of multimodal data, such as combining imaging data with genomic and clinical information to provide models with richer information that enhances diagnostic accuracy.
- Investigating explainable AI methods to improve transparency and interpretability of model classifications, thus promoting greater trust and acceptance among clinicians.
- Developing user-friendly interfaces and tools to ensure the smooth integration of these models into clinical practice, enhancing their effectiveness in breast cancer care.

6.4 Publications

6.4.1 Journals

Beghriche, T., Attallah, B., Brik, Y., & Djerioui, M. (2023). A multi-level fine-tuned deep learning based approach for binary classification of diabetic retinopathy. *Chemometrics and Intelligent Laboratory Systems*, 237, 104820.

Heythem, B., Djerioui, M., **Beghriche, T.**, Zerguine, A., & Beghdadi, A. (2024). Customized CNN for Multi-Class Classification of Brain Tumor Based on MRI Images. *Arabian Journal for Science and Engineering*, 1–16.

Beghriche, T., Brik, Y., Djerioui, M., Attallah, B., Zerguine, A., & Beghdadi, A. (2025). A Multi-stage Optimization Architecture for Effective Breast Cancer Diagnosis Based on Deep Neural Networks. *Arabian Journal for Science and Engineering*, 1–26.

6.4.2 International Conferences

Beghriche, T., Djerioui, M., & Brik, Y. (2022, November). Performance Analysis of Twin-Support Vector Machine in Breast Cancer Prediction. In *2022 International Conference of Advanced Technology in Electronic and Electrical Engineering (ICATEEE)* (pp. 1–5). IEEE.

Beghriche, T., Djerioui, M., & Brik, Y. (15–16 June 2024). Enhancing Diabetic Retinopathy Detection Using a Hybrid Framework Integrating Machine Learning Classifiers and Advanced Feature Extraction Techniques. *The 3rd International Conference on Frontiers in Academic Research*, Konya-Turkey.

6.4.3 National Conferences

Beghriche, T., Djerioui, M., & Brik, Y. (17–18 November 2024). An innovative CNN-SVM hybrid model for enhanced diabetic retinopathy detection. *The 8th National Colloquium on Inductics: University-Industry*, M'Sila-Algeria.

Beghriche, T., Djerioui, M., Brik, Y. & Attallah, B. (17–18 November 2024). Enhancing diabetic retinopathy classification with vision transformers (ViTs). *The 8th National Colloquium on Inductics: University-Industry*, M'Sila-Algeria.

Beghriche, T., Djerioui, M., Brik, Y. & Attallah, B. (17 December 2024). A Hybrid Framework for Enhanced Diabetic Retinopathy Detection Using Advanced Deep CNN-Based Feature

Extractors and Machine Learning Classifiers. *The Third National Conference on Mathematics, Biology, and Medicine*, M'sila-Algeria.

Beghriche, T., Djerioui, M., & Brik, Y. (22–23 April 2025). A voting ensemble machine learning-based method for improved breast cancer classification. *The First National Conference on Emerging Trends in Engineering and Technology (NC2ET'25)*, Mascara-Algeria.

Bibliography

[1] Adyasha Sahu, Pradeep Kumar Das, and Sukadev Meher. High accuracy hybrid cnn classifiers for breast cancer detection using mammogram and ultrasound datasets. *Biomedical Signal Processing and Control*, 80:104292, 2023.

[2] World Health Organization. Breast Cancer Fact Sheet. https://www.who.int/news-room/fact-sheets/detail/breast-cancer. Accessed: 2022-11-06.

[3] Mina Samieinasab, S Ahmad Torabzadeh, Arman Behnam, Amir Aghsami, and Fariborz Jolai. Meta-health stack: A new approach for breast cancer prediction. *Healthcare Analytics*, 2:100010, 2022.

[4] Riyadh M Al-Tam and Sachin M Narangale. Breast cancer detection and diagnosis using machine learning: a survey. *J. Sci. Res*, 65(5):265–285, 2021.

[5] Min Jung Ko, Dong A Park, Sung Hyun Kim, Eun Sook Ko, Kyung Hwan Shin, Woosung Lim, Beom Seok Kwak, and Jung Min Chang. Accuracy of digital breast tomosynthesis for detecting breast cancer in the diagnostic setting: a systematic review and meta-analysis. *Korean journal of radiology*, 22(8):1240, 2021.

[6] Sven CA Michel, Thomas M Keller, Johannes M Fr"ohlich, Daniel Fink, Rosmarie Caduff, Burkhardt Seifert, Borut Marincek, and Rahel A Kubik-Huch. Preoperative breast cancer staging: Mr imaging of the axilla with ultrasmall superparamagnetic iron oxide enhancement. *Radiology*, 225(2):527–536, 2002.

[7] James V Fiorica. Breast cancer screening, mammography, and other modalities. *Clinical obstetrics and gynecology*, 59(4):688–709, 2016.

[8] Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. A guide to deep learning in healthcare. *Nature medicine*, 25(1):24–29, 2019.

[9] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[10] Mehdi Habibzadeh Motlagh, Mahboobeh Jannesari, HamidReza Aboulkheyr, Pegah Khosravi, Olivier Elemento, Mehdi Totonchi, and Iman Hajirasouliha. Breast cancer histopathological image classification: A deep learning approach. *BioRxiv*, page 242818, 2018.

[11] Scott Mayer McKinney, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, Hutan Ashrafian, Trevor Back, Mary Chesus, Greg S Corrado, Ara Darzi, et al. International evaluation of an ai system for breast cancer screening. *Nature*, 577(7788):89–94, 2020.

[12] Berkman Sahiner, Aria Pezeshk, Lubomir M Hadjiiski, Xiaosong Wang, Karen Drukker, Kenny H Cha, Ronald M Summers, and Maryellen L Giger. Deep learning in medical imaging and radiation therapy. *Medical physics*, 46(1):e1–e36, 2019.

[13] Dezső Ribli, Anna Horváth, Zsuzsa Unger, Péter Pollner, and István Csabai. Detecting and classifying lesions in mammograms with deep learning. *Scientific reports*, 8(1):4165, 2018.

[14] Constance D Lehman, Robert D Wellman, Diana SM Buist, Karla Kerlikowske, Anna NA Tosteson, Diana L Miglioretti, Breast Cancer Surveillance Consortium, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA internal medicine*, 175(11):1828–1837, 2015.

[15] Nathalie Japkowicz and Shaju Stephen. The class imbalance problem: A systematic study. *Intelligent data analysis*, 6(5):429–449, 2002.

[16] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[17] Junjun Zhang, Rosita Bajari, Dusan Andric, Francois Gerthoffert, Alexandru Lepsa, Hardeep Nahal-Bose, Lincoln D Stein, and Vincent Ferretti. The international cancer genome consortium data portal. *Nature biotechnology*, 37(4):367–369, 2019.

[18] Vivek Kumar, Brojo Kishore Mishra, Manuel Mazzara, Dang NH Thanh, and Abhishek Verma. Prediction of malignant and benign breast cancer: A data mining approach in healthcare applications. In *Advances in Data Science and Management: Proceedings of ICDSM 2019*, pages 435–442. Springer, 2020.

[19] Rayees Ahmad Dar, Muzafar Rasool, Assif Assad, et al. Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. *Computers in biology and medicine*, 149:106073, 2022.

[20] Jer-Wei Chang, Wen-Hung Kuo, Chiao-Mei Lin, Wen-Ling Chen, Shih-Hsuan Chan, Meng-Fan Chiu, I-Shou Chang, Shih-Sheng Jiang, Fang-Yu Tsai, Chung-Hsing Chen, et al. Wild-type p53 upregulates an early onset breast cancer-associated gene gas7 to suppress metastasis via gas7–cyfip1-mediated signaling pathway. *Oncogene*, 37(30):4137–4150, 2018.

[21] M. Zwitter and Matjaz Soklic. Breast cancer. https://archive-beta.ics.uci.edu/ml/datasets/breast+cancer, 1988. Accessed: 2022-05-06.

[22] Na Liu, Er-Shi Qi, Man Xu, Bo Gao, and Gui-Qiu Liu. A novel intelligent classification model for breast cancer diagnosis. *Information Processing & Management*, 56(3):609–623, 2019.

[23] Ziba Khandezamin, Marjan Naderan, and Mohammad Javad Rashti. Detection and classification of breast cancer using logistic regression feature selection and gmdh classifier. *Journal of Biomedical Informatics*, 111:103591, 2020.

[24] Deepti Sharma, Rajneesh Kumar, and Anurag Jain. Breast cancer prediction based on neural networks and extra tree classifier using feature ensemble learning. *Measurement: Sensors*, 24:100560, 2022.

[25] Zexian Huang and Daqi Chen. A breast cancer diagnosis method based on vim feature selection and hierarchical clustering random forest algorithm. *IEEE Access*, 10:3284–3293, 2021.

[26] Law Kumar Singh, Munish Khanna, and Rekha Singh. Artificial intelligence based medical decision support system for early and accurate breast cancer prediction. *Advances in engineering software*, 175:103338, 2023.

[27] Mehdi Khashei and Negar Bakhtiarvand. A novel discrete learning-based intelligent methodology for breast cancer classification purposes. *Artificial Intelligence In Medicine*, 139:102492, 2023.

[28] Muhammad Sakib Khan Inan, Sohrab Hossain, and Mohammed Nazim Uddin. Data augmentation guided breast cancer diagnosis and prognosis using an integrated deep-generative framework based on breast tumor's morphological information. *Informatics in Medicine Unlocked*, 37:101171, 2023.

[29] Sutong Wang, Yuyan Wang, Dujuan Wang, Yunqiang Yin, Yanzhang Wang, and Yaochu Jin. An improved random forest-based rule extraction method for breast cancer diagnosis. *Applied Soft Computing*, 86:105941, 2020.

[30] Pratheep Kumar, Geetha G Nair, et al. An efficient classification framework for breast cancer using hyper parameter tuned random decision forest classifier and bayesian optimization. *Biomedical Signal Processing and Control*, 68:102682, 2021.

[31] Amr E Eldin Rashed, Ashraf M Elmorsy, and Ahmed E Mansour Atwa. Comparative evaluation of automated machine learning techniques for breast cancer diagnosis. *Biomedical Signal Processing and Control*, 86:105016, 2023.

[32] Ramdas Kapila and Sumalatha Saleti. An efficient ensemble-based machine learning for breast cancer detection. *Biomedical Signal Processing and Control*, 86:105269, 2023.

[33] Saad Almutairi, S Manimurugan, Byung-Gyu Kim, Majed M Aborokbah, and C Narmatha. Breast cancer classification using deep q learning (dql) and gorilla troops optimization (gto). *Applied Soft Computing*, 142:110292, 2023.

[34] Emilija Strelcenia and Simant Prakoonwit. Improving cancer detection classification performance using gans in breast cancer data. *IEEE Access*, 2023.

[35] Marion Olubunmi Adebiyi, Micheal Olaolu Arowolo, Moses Damilola Mshelia, and Oludayo O Olugbara. A linear discriminant analysis and classification model for breast cancer diagnosis. *Applied Sciences*, 12(22):11455, 2022.

[36] AS Vibith and Jobin Christ. Gbdtmo: as new option for early-stage breast cancer detection and classification using machine learning. *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, 64(4):858–867, 2023.

[37] Fatemeh Hamedani-KarAzmoudehFar, Reza Tavakkoli-Moghaddam, Amir Reza Tajally, and Seyed Sina Aria. Breast cancer classification by a new approach to assessing deep neural network-based uncertainty quantification methods. *Biomedical Signal Processing and Control*, 79:104057, 2023.

[38] ON Oyelade, AA Obiniyi, SB Junaidu, and SA Adewuyi. St-oncodiag: A semantic rule-base approach to diagnosing breast cancer base on wisconsin datasets. *Informatics in Medicine Unlocked*, 10:117–125, 2018.

[39] Muhammad Hammad Memon, Jian Ping Li, Amin Ul Haq, Muhammad Hunain Memon, and Wang Zhou. Breast cancer detection in the iot health environment using modified recursive feature selection. *wireless communications and mobile computing*, 2019(1):5176705, 2019.

[40] Elsayed Badr, Sultan Almotairi, Mustafa Abdul Salam, and Hagar Ahmed. New sequential and parallel support vector machine with grey wolf optimizer for breast cancer diagnosis. *Alexandria Engineering Journal*, 61(3):2520–2534, 2022.

[41] O Ibrahim Obaid, Mazin Abed Mohammed, Mohd Kanapi Abd Ghani, A Mostafa, Fahad Taha, et al. Evaluating the performance of machine learning techniques in the classification of wisconsin breast cancer. *International Journal of Engineering & Technology*, 7(4.36):160–166, 2018.

[42] Ahmed S Elkorany, Mohamed Marey, Khaled M Almustafa, and Zeinab F Elsharkawy. Breast cancer diagnosis using support vector machines optimized by whale optimization and dragonfly algorithms. *IEEE Access*, 10:69688–69699, 2022.

[43] Muhammet Fatih Ak. A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications. In *Healthcare*, volume 8, page 111. MDPI, 2020.

[44] Nosayba Al-Azzam and Ibrahem Shatnawi. Comparing supervised and semi-supervised machine learning models on diagnosing breast cancer. *Annals of Medicine and Surgery*, 62:53–64, 2021.

[45] WH Wolberg and M Olvi. Breast cancer wisconsin (original) data set, uci machine learning repository, 1992.

[46] Hamideh Hajiabadi, Vahide Babaiyan, Davood Zabihzadeh, and Moein Hajiabadi. Combination of loss functions for robust breast cancer prediction. *Computers & Electrical Engineering*, 84:106624, 2020.

[47] Sawssen Bacha and Okba Taouali. A novel machine learning approach for breast cancer diagnosis. *Measurement*, 187:110233, 2022.

[48] V Nanda Gopal, Fadi Al-Turjman, R Kumar, L Anand, and M Rajesh. Feature selection and classification in breast cancer prediction using iot and machine learning. *Measurement*, 178:109442, 2021.

[49] Moloud Abdar and Vladimir Makarenkov. Cwv-bann-svm ensemble learning classifier for an accurate diagnosis of breast cancer. *Measurement*, 146:557–570, 2019.

[50] Yash Amethiya, Prince Pipariya, Shlok Patel, and Manan Shah. Comparative analysis of breast cancer detection using machine learning and biosensors. *Intelligent Medicine*, 2(2):69–81, 2022.

[51] Amreen Batool and Yung-Cheol Byun. Towards improving breast cancer classification using an adaptive voting ensemble learning algorithm. *IEEE Access*, 2024.

[52] Chour Singh Rajpoot, Gajanand Sharma, Praveen Gupta, Pankaj Dadheech, Umar Yahya, and Nagender Aneja. Feature selection-based machine learning comparative analysis for predicting breast cancer. *Applied Artificial Intelligence*, 38(1):2340386, 2024.

[53] Zhiqiang Guo, Lina Xu, and Nona Ali Asgharzadeholiaee. A homogeneous ensemble classifier for breast cancer detection using parameters tuning of mlp neural network. *Applied Artificial Intelligence*, 36(1):2031820, 2022.

[54] Long Jin, Zhiguan Huang, Liangming Chen, Mei Liu, Yuhe Li, Yao Chou, and Chenfu Yi. Modified single-output chebyshev-polynomial feedforward neural network aided with subset method for classification of breast cancer. *Neurocomputing*, 350:128–135, 2019.

[55] Ahmed Hamza Osman and Hani Moetque Abdullah Aljahdali. An effective of ensemble boosting learning method for breast cancer virtual screening using neural network model. *IEEE Access*, 8:39165–39174, 2020.

[56] Vikas Chaurasia, Saurabh Pal, and BB Tiwari. Prediction of benign and malignant breast cancer using data mining techniques. *Journal of Algorithms & Computational Technology*, 12(2):119–126, 2018.

[57] William Wolberg, W Street, and Olvi Mangasarian. Breast cancer wisconsin (prognostic). uci machine learning repository, 1995.

[58] Maram Alwohaibi, Malek Alzaqebah, Noura M Alotaibi, Abeer M Alzahrani, and Mariem Zouch. A hybrid multi-stage learning technique based on brain storming optimization algorithm for breast cancer recurrence prediction. *Journal of King Saud University-Computer and Information Sciences*, 34(8):5192–5203, 2022.

[59] UCI Machine Learning Repository. Uci machine learning repository. https://archive-beta.ics.uci.edu/ml/datasets/breast+cancer+coimbra, 2018. Accessed: 2022-05-06.

[60] S Nanglia, Muneer Ahmad, Fawad Ali Khan, and NZ Jhanjhi. An enhanced predictive heterogeneous ensemble model for breast cancer prediction. *Biomedical Signal Processing and Control*, 72:103279, 2022.

[61] Rolando Gonzales Martinez and Daan-Max van Dongen. Deep learning algorithms for the early detection of breast cancer: A comparative study with traditional machine learning. *Informatics in Medicine Unlocked*, 41:101317, 2023.

[62] J. Teng. Seer breast cancer data. IEEE DataPort, 2019. Accessed: 2022-05-07.

[63] Yafei Wu, Yaheng Zhang, Siyu Duan, Chenming Gu, Chongtao Wei, and Ya Fang. Survival prediction in second primary breast cancer patients with machine learning: An analysis of seer database. *Computer Methods and Programs in Biomedicine*, 254:108310, 2024.

[64] METABRIC Consortium. Metabric genomics dataset. The European genome-phenome archive (EGA). Accessed: May 07, 2022.

[65] Parampreet Kaur, Ashima Singh, and Inderveer Chana. Bsense: A parallel bayesian hyperparameter optimized stacked ensemble model for breast cancer survival prediction. *Journal of Computational Science*, 60:101570, 2022.

[66] Gaetano Manzo, Yann Pannatier, Pascal Duflot, et al. Cohort and trajectory analysis in multi-agent support systems for cancer survivors. *Computer Methods and Programs in Biomedicine*, 231:107373, 2023.

[67] Nikhilanand Arya and Sriparna Saha. Multi-modal advanced deep learning architectures for breast cancer survival prediction. *Knowledge-Based Systems*, 221:106965, 2021.

[68] Nikhilanand Arya and Sriparna Saha. Multi-modal classification for human breast cancer prognosis prediction: proposal of deep-learning based stacked ensemble model. *IEEE/ACM transactions on computational biology and bioinformatics*, 19(2):1032–1041, 2020.

[69] Nashwa El-Bendary, Nahla A Belal, et al. A feature-fusion framework of clinical, genomics, and histopathological data for metabric breast cancer subtype classification. *Applied Soft Computing*, 91:106238, 2020.

[70] Nermin Abdelhakim Othman, Manal A Abdel-Fattah, and Ahlam Talaat Ali. A hybrid deep learning framework with decision-level fusion for breast cancer survival prediction. *Big Data and Cognitive Computing*, 7(1):50, 2023.

[71] Haozhe Xie, Jie Li, Qiaosheng Zhang, and Yadong Wang. Comparison among dimensionality reduction techniques based on random projection for cancer classification. *Computational biology and chemistry*, 65:165–172, 2016.

[72] V Nandagopal, S Geeitha, K Vinoth Kumar, and J Anbarasi. Feasible analysis of gene expression–a computational based classification for breast cancer. *Measurement*, 140:120–125, 2019.

[73] Jiande Wu and Chindo Hicks. Breast cancer type classification using machine learning. *Journal of personalized medicine*, 11(2):61, 2021.

[74] Abhishek Das, Mihir Narayan Mohanty, Pradeep Kumar Mallick, Prayag Tiwari, Khan Muhammad, and Hongyin Zhu. Breast cancer detection using an ensemble deep learning method. *Biomedical Signal Processing and Control*, 70:103009, 2021.

[75] Hanaa Torkey, Mostafa Atlam, Nawal El-Fishawy, and Hanaa Salem. A novel deep autoencoder based survival analysis approach for microarray dataset. *PeerJ Computer Science*, 7:e492, 2021.

[76] Wenlong Ming, Fuyu Li, Yanhui Zhu, Yunfei Bai, Wanjun Gu, Yun Liu, Xiao Sun, Xiaoan Liu, and Hongde Liu. Predicting hormone receptors and pam50 subtypes of breast cancer from multi-scale lesion images of dce-mri with transfer learning technique. *Computers in biology and medicine*, 150:106147, 2022.

[77] Jinshan Tang, Rangaraj M Rangayyan, Jun Xu, Issam El Naqa, and Yongyi Yang. Computer-aided detection and diagnosis of breast cancer with mammography: recent

advances. *IEEE transactions on information technology in biomedicine*, 13(2):236–251, 2009.

[78] Michael Heath, Kevin Bowyer, Daniel Kopans, P Kegelmeyer Jr, Richard Moore, Kyong Chang, and S Munishkumaran. Current status of the digital database for screening mammography. In *Digital Mammography: Nijmegen, 1998*, pages 457–460. Springer, 1998.

[79] Lal Hussain, Shahzad Ahmad Qureshi, Amjad Aldweesh, Jawad ur Rehman Pirzada, Faisal Mehmood Butt, Elsayed Tag Eldin, Mushtaq Ali, Abdulmohsen Algarni, and Muhammad Amin Nadim. Automated breast cancer detection by reconstruction independent component analysis (rica) based hybrid features using machine learning paradigms. *Connection Science*, 34(1):2784–2806, 2022.

[80] Debendra Muduli, Ratnakar Dash, and Banshidhar Majhi. Fast discrete curvelet transform and modified pso based improved evolutionary extreme learning machine for breast cancer detection. *Biomedical Signal Processing and Control*, 70:102919, 2021.

[81] Mugahed A Al-Antari, Seung-Moo Han, and Tae-Seong Kim. Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital x-ray mammograms. *Computer methods and programs in biomedicine*, 196:105584, 2020.

[82] Ayşe Aydın Yurdusev, Kemal Adem, and Mahmut Hekim. Detection and classification of microcalcifications in mammograms images using difference filter and yolov4 deep learning model. *Biomedical Signal Processing and Control*, 80:104360, 2023.

[83] Khaoula Belhaj Soulami, Naima Kaabouch, and Mohamed Nabil Saidi. Breast cancer: Classification of suspicious regions in digital mammograms based on capsule network. *Biomedical Signal Processing and Control*, 76:103696, 2022.

[84] S Vidivelli and S Sathiya Devi. Breast cancer detection model using fuzzy entropy segmentation and ensemble classification. *Biomedical Signal Processing and Control*, 80:104236, 2023.

[85] Maria Wimmer, Gert Sluiter, David Major, Dimitrios Lenis, Astrid Berg, Theresa Neubauer, and Katja Bühler. Multi-task fusion for improving mammography screening data classification. *IEEE Transactions on Medical Imaging*, 41(4):937–950, 2021.

[86] R Sathesh Raaj. Breast cancer detection and diagnosis using hybrid deep learning architecture. *Biomedical Signal Processing and Control*, 82:104558, 2023.

[87] A Abdul Hayum, J Jaya, B Paulchamy, and R Sivakumar. A modified recurrent neural network (mrnn) model for and breast cancer classification system. *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, 64(4):1193–1203, 2023.

[88] Mohaddeseh Chegini and Ali Mahlooji Far. Uncertainty-aware deep learning-based cad system for breast cancer classification using ultrasound and mammography images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 12(1):2297983, 2024.

[89] Jawad Ahmad, Sheeraz Akram, Arfan Jaffar, Muhammad Rashid, and Sohail Masood Bhatti. Breast cancer detection using deep learning: An investigation using the ddsm dataset and a customized alexnet and support vector machine. *IEEE Access*, 2023.

[90] Sarmad Maqsood, Robertas Damaševičius, and Rytis Maskeliūnas. Ttcnn: A breast cancer detection and classification towards computer-aided diagnosis using digital mammography in early stages. *Applied Sciences*, 12(7):3273, 2022.

[91] Sathish Kumar PJ, S Shibu, M Mohan, and T Kalaichelvi. Hybrid deep learning enabled breast cancer detection using mammogram images. *Biomedical Signal Processing and Control*, 95:106310, 2024.

[92] Rebecca Sawyer Lee, Francisco Gimenez, Assaf Hoogi, Kanae Kawai Miyake, Mia Gorovoy, and Daniel L Rubin. A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific data*, 4(1):1–9, 2017.

[93] Alessia Gerbasi, Greta Clementi, Fabio Corsi, Sara Albasini, Alberto Malovini, Silvana Quaglini, and Riccardo Bellazzi. Deepmica: Automatic segmentation and classification of breast microcalcifications from mammograms. *Computer Methods and Programs in Biomedicine*, 235:107483, 2023.

[94] Hiba Chougrad, Hamid Zouaki, and Omar Alheyane. Multi-label transfer learning for the early diagnosis of breast cancer. *Neurocomputing*, 392:168–180, 2020.

[95] Jiale Jiang, Junchuan Peng, Chuting Hu, Wenjing Jian, Xianming Wang, and Weixiang Liu. Breast cancer detection and classification in mammogram using a three-stage deep learning framework based on paa algorithm. *Artificial Intelligence in Medicine*, 134:102419, 2022.

[96] Ahmed Iqbal and Muhammad Sharif. Bts-st: Swin transformer network for segmentation and classification of multimodality breast cancer images. *Knowledge-Based Systems*, 267:110393, 2023.

[97] Volkan Muejdat Tiryaki. Mass segmentation and classification from film mammograms using cascaded deep transfer learning. *Biomedical Signal Processing and Control*, 84:104819, 2023.

[98] Lydia Bouzar-Benlabiod, Khaled Harrar, Lahcen Yamoun, Mustapha Yacine Khodja, and Moulay A Akhloufi. A novel breast cancer detection architecture based on a cnn-cbr system for mammogram classification. *Computers in biology and medicine*, 163:107133, 2023.

[99] Savita Kumbhare, Atul B Kathole, and Swati Shinde. Federated learning aided breast cancer detection with intelligent heuristic-based deep learning framework. *Biomedical Signal Processing and Control*, 86:105080, 2023.

[100] Unaiza Sajid, Rizwan Ahmed Khan, Shahid Munir Shah, and Sheeraz Arif. Breast cancer classification using deep learned features boosted with handcrafted features. *Biomedical Signal Processing and Control*, 86:105353, 2023.

[101] Leren Qian, Jiexin Bai, Yiqian Huang, Diyar Qader Zeebaree, Abbas Saffari, and Dilovan Asaad Zebari. Breast cancer diagnosis using evolving deep convolutional neural network based on hybrid extreme learning machine technique and improved chimp optimization algorithm. *Biomedical Signal Processing and Control*, 87:105492, 2024.

[102] Inês C Moreira, Igor Amaral, Inês Domingues, António Cardoso, Maria Joao Cardoso, and Jaime S Cardoso. Inbreast: toward a full-field digital mammographic database. *Academic radiology*, 19(2):236–248, 2012.

[103] SR Sannasi Chakravarthy, N Bharanidharan, and H Rajaguru. Deep learning-based metaheuristic weighted k-nearest neighbor algorithm for the severity classification of breast cancer. *IRBM*, 44(3):100749, 2023.

[104] Bita Asadi and Qurban Memon. Efficient breast cancer detection via cascade deep learning network. *International Journal of Intelligent Networks*, 4:46–52, 2023.

[105] Muhammet Fatih Aslan. A hybrid end-to-end learning approach for breast cancer diagnosis: convolutional recurrent network. *Computers and Electrical Engineering*, 105:108562, 2023.

[106] R Karthiga, Kumaravelu Narasimhan, and Rengarajan Amirtharajan. Diagnosis of breast cancer for modern mammography using artificial intelligence. *Mathematics and Computers in Simulation*, 202:316–330, 2022.

[107] Amelia Jiménez-Sánchez, Mickael Tardy, Miguel A González Ballester, Diana Mateus, and Gemma Piella. Memory-aware curriculum federated learning for breast cancer classification. *Computer Methods and Programs in Biomedicine*, 229:107318, 2023.

[108] John Suckling. The mammographic images analysis society digital mammogram database. In *Exerpta Medica. International Congress Series, 1994*, volume 1069, pages 375–378, 1994.

[109] Smita Khairnar, Sudeep D Thepade, and Shilpa Gite. Effect of image binarization thresholds on breast cancer identification in mammography images using otsu, niblack, burnsen, thepade's sbtc. *Intelligent Systems with Applications*, 10:200046, 2021.

[110] Debendra Muduli, Ratnakar Dash, and Banshidhar Majhi. Automated breast cancer detection in digital mammograms: A moth flame optimization based elm approach. *Biomedical Signal Processing and Control*, 59:101912, 2020.

[111] M Saeed Darweesh, Mostafa Adel, Ahmed Anwar, Omar Farag, Ahmed Kotb, Mohamed Adel, Ayman Tawfik, and Hassan Mostafa. Early breast cancer diagnostics based on hierarchical machine learning classification for mammography images. *Cogent Engineering*, 8(1):1968324, 2021.

[112] Kevin M Kelly, Judy Dean, W Scott Comulada, and Sung-Jae Lee. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *European radiology*, 20:734–742, 2010.

[113] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in brief*, 28:104863, 2020.

[114] Iulia-Nela Anghelache Nastase, Simona Moldovanu, and Luminita Moraru. Image moment-based features for mass detection in breast us images via machine learning and neural network classification models. *Inventions*, 7(2):42, 2022.

[115] Si-Yuan Lu, Shui-Hua Wang, and Yu-Dong Zhang. Safnet: A deep spatial attention network with classifier fusion for breast cancer detection. *Computers in Biology and Medicine*, 148:105812, 2022.

[116] S-Y Lu, S-H Wang, and Y-D Zhang. Bcdnet: an optimized deep network for ultrasound breast cancer detection. *IRBM*, 44(4):100774, 2023.

[117] Junxia Wang, Yuanjie Zheng, Jun Ma, Xinmeng Li, Chongjing Wang, James Gee, Haipeng Wang, and Wenhui Huang. Information bottleneck-based interpretable multitask network for breast cancer classification and segmentation. *Medical image analysis*, 83:102687, 2023.

[118] Sagar Deep Deb and Rajib Kumar Jha. Breast ultrasound image classification using fuzzy-rank-based ensemble network. *Biomedical Signal Processing and Control*, 85:104871, 2023.

[119] Alessandro Sebastian Podda, Riccardo Balia, Silvio Barra, Salvatore Carta, Gianni Fenu, and Leonardo Piano. Fully-automated deep learning pipeline for segmentation and classification of breast ultrasound images. *Journal of computational science*, 63:101816, 2022.

[120] Muhammad Sakib Khan Inan, Fahim Irfan Alam, and Rizwan Hasan. Deep integrated pipeline of segmentation guided classification of breast cancer from ultrasound images. *Biomedical Signal Processing and Control*, 75:103553, 2022.

[121] Madhusudan G Lanjewar, Kamini G Panchbhai, and Lalchand B Patle. Fusion of transfer learning models with lstm for detection of breast cancer using ultrasound images. *Computers in Biology and Medicine*, 169:107914, 2024.

[122] Mohsin Furkh Dar and Avatharam Ganivada. Deep learning and genetic algorithm-based ensemble model for feature selection and classification of breast ultrasound images. *Image and Vision Computing*, 146:105018, 2024.

[123] Haixia Liu, Guozhong Cui, Yi Luo, Yajie Guo, Lianli Zhao, Yueheng Wang, Abdulhamit Subasi, Sengul Dogan, and Turker Tuncer. Artificial intelligence-based breast cancer diagnosis using ultrasound images and grid-based deep feature generator. *International Journal of General Medicine*, pages 2271–2282, 2022.

[124] Boyu Zhang, Aleksandar Vakanski, and Min Xian. Bi-rads-net-v2: A composite multi-task neural network for computer-aided diagnosis of breast cancer in ultrasound images with semantic and quantitative explanations. *IEEE Access*, 2023.

[125] Kuncham Sreenivasa Rao, Panduranga Vital Terlapu, D Jayaram, K Kishore Raju, G Kiran Kumar, Rambabu Pemula, Venu Gopalachari, and S Rakesh. Intelligent ultrasound imaging for enhanced breast cancer diagnosis: Ensemble transfer learning strategies. *IEEE Access*, 2024.

[126] Kiran Jabeen, Muhammad Attique Khan, Majed Alhaisoni, Usman Tariq, Yu-Dong Zhang, Ameer Hamza, Artūras Mickus, and Robertas Damaševičius. Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion. *Sensors*, 22(3):807, 2022.

[127] Marwa Obayya, Siwar Ben Haj Hassine, Sana Alazwari, Mohamed K. Nour, Abdullah Mohamed, Abdelwahed Motwakel, Ishfaq Yaseen, Abu Sarwar Zamani, Amgad Atta Abdelmageed, and Gouse Pasha Mohammed. Aquila optimizer with bayesian neural network for breast cancer detection on ultrasound images. *Applied Sciences*, 12(17):8679, 2022.

[128] Asaf Raza, Naeem Ullah, Javed Ali Khan, Muhammad Assam, Antonella Guzzo, and Hanan Aljuaid. Deepbreastcancernet: A novel deep learning model for breast cancer detection using ultrasound images. *Applied Sciences*, 13(4):2082, 2023.

[129] Qiqi He, Qiuju Yang, Hang Su, and Yixuan Wang. Multi-task learning for segmentation and classification of breast tumors from ultrasound images. *Computers in Biology and Medicine*, 173:108319, 2024.

[130] Adyasha Sahu, Pradeep Kumar Das, and Sukadev Meher. An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images. *Biomedical Signal Processing and Control*, 87:105377, 2024.

[131] Chengzhang Zhu, Xian Chai, Zhiyuan Wang, Yalong Xiao, RenMao Zhang, Zhangzheng Yang, and Jie Feng. Dbl-net: A dual-branch learning network with information from spatial and frequency domains for tumor segmentation and classification in breast ultrasound image. *Biomedical Signal Processing and Control*, 93:106221, 2024.

[132] Moi Hoon Yap, Gerard Pons, Joan Marti, Sergi Ganau, Melcior Sentis, Reyer Zwiggelaar, Adrian K Davison, and Robert Marti. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE journal of biomedical and health informatics*, 22(4):1218–1226, 2017.

[133] Dezhuang Kong, Shunbo Hu, and Guojia Zhao. Mv-stcnet: Breast cancer diagnosis using spatial and temporal dual-attention guided classification network based on multi-view ultrasound videos. *Biomedical Signal Processing and Control*, 87:105541, 2024.

[134] Dinghao Guo, Chunyu Lu, Dali Chen, Jizhong Yuan, Qimu Duan, Zheng Xue, Shixin Liu, and Ying Huang. A multimodal breast cancer diagnosis method based on knowledge-augmented deep learning. *Biomedical Signal Processing and Control*, 90:105843, 2024.

[135] Mireille Van Goethem, W Tjalma, K Schelfout, I Verslegers, I Biltjes, and P Parizel. Magnetic resonance imaging in breast cancer. *European Journal of Surgical Oncology (EJSO)*, 32(9):901–910, 2006.

[136] Doris Leithner, Marius E Mayerhoefer, Danny F Martinez, Maxine S Jochelson, Elizabeth A Morris, Sunitha B Thakur, and Katja Pinker. Non-invasive assessment of breast cancer molecular subtypes with multiparametric magnetic resonance imaging radiomics. *Journal of clinical medicine*, 9(6):1853, 2020.

[137] Cancer Imaging Archive. Breast mri nact pilot. https://wiki.cancerimagingarchive.net/display/Public/Breast-MRI-NACT-Pilot. Accessed: 2024-10-28.

[138] Zhe Zhu, Ehab Albadawy, Ashirbani Saha, Jun Zhang, Michael R Harowicz, and Maciej A Mazurowski. Deep learning for identifying radiogenomic associations in breast cancer. *Computers in biology and medicine*, 109:85–90, 2019.

[139] Priyanka Khanna, Mridu Sahu, Bikesh Kumar Singh, and Vikrant Bhateja. Early prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer mri images using combined pre-trained convolutional neural network and machine learning. *Measurement*, 207:112269, 2023.

[140] Cancer Imaging Archive. Data from rider-breast-mri. https://wiki.cancerimagingarchive.net/display/Public/RIDER+Breast+MRI, 2015. Accessed: 2024-10-28.

[141] W Lingle, BJ Erickson, ML Zuley, R Jarosz, E Bonaccio, J Filippini, and N Gruszauskas. Radiology data from the cancer genome atlas breast invasive carcinoma [tcga-brca] collection. *The Cancer Imaging Archive*, 10(K9):5, 2016.

[142] T Liu, J Huang, T Liao, R Pu, S Liu, and Y Peng. A hybrid deep learning model for predicting molecular subtypes of human breast cancer using multimodal data. *Irbm*, 43(1):62–74, 2022.

[143] Yoni Schirris, Efstratios Gavves, Iris Nederlof, Hugo Mark Horlings, and Jonas Teuwen. Deepsmile: Contrastive self-supervised pre-training benefits msi and hrd classification directly from h&e whole-slide images in colorectal and breast cancer. *Medical Image Analysis*, 79:102464, 2022.

[144] Metin N Gurcan, Laura E Boucheron, Ali Can, Anant Madabhushi, Nasir M Rajpoot, and Bulent Yener. Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, 2:147–171, 2009.

[145] Fabio A Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte. A dataset for breast cancer histopathological image classification. *Ieee transactions on biomedical engineering*, 63(7):1455–1462, 2015.

[146] Usman Haider, Muhammad Hanif, Ahmar Rashid, Khursheed Aurangzeb, Akhtar Khalil, and Musaed Alhussein. Discriminative dictionary learning using penalized rank-1 approximation for breast cancer classification with imbalanced dataset. *IEEE Access*, 2023.

[147] Dan Wang, Zhen Chen, and Hongwei Zhao. Prototype transfer generative adversarial network for unsupervised breast cancer histology image classification. *Biomedical Signal Processing and Control*, 68:102713, 2021.

[148] Agaba Ameh Joseph, Mohammed Abdullahi, Sahalu Balarabe Junaidu, Hayatu Hassan Ibrahim, and Haruna Chiroma. Improved multi-classification of breast cancer histopathological images using handcrafted features and deep neural network (dense layer). *Intelligent Systems with Applications*, 14:200066, 2022.

[149] Mohammad Amin Shamshiri, Adam Krzyżak, Marek Kowal, and Józef Korbicz. Compatible-domain transfer learning for breast cancer classification with limited annotated data. *Computers in Biology and Medicine*, 154:106575, 2023.

[150] Mohammad Reza Abbasniya, Sayed Ali Sheikholeslamzadeh, Hamid Nasiri, and Samaneh Emami. Classification of breast tumors based on histopathology images using deep features and ensemble of gradient boosting methods. *Computers and Electrical Engineering*, 103:108382, 2022.

[151] Ramgopal Kashyap. Dilated residual grooming kernel model for breast cancer detection. *Pattern Recognition Letters*, 159:157–164, 2022.

[152] Hanan Aljuaid, Nazik Alturki, Najah Alsubaie, Lucia Cavallaro, and Antonio Liotta. Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. *Computer Methods and Programs in Biomedicine*, 223:106951, 2022.

[153] Said Boumaraf, Xiabi Liu, Zhongshu Zheng, Xiaohong Ma, and Chokri Ferkous. A new transfer learning based approach to magnification dependent and independent classifi-

cation of breast cancer in histopathological images. *Biomedical Signal Processing and Control*, 63:102192, 2021.

[154] Soham Chattopadhyay, Arijit Dey, Pawan Kumar Singh, Diego Oliva, Erik Cuevas, and Ram Sarkar. Mtrre-net: A deep learning model for detection of breast cancer from histopathological images. *Computers in Biology and Medicine*, 150:106155, 2022.

[155] Zhanbo Yang, Lingyan Ran, Shizhou Zhang, Yong Xia, and Yanning Zhang. Ems-net: Ensemble of multiscale convolutional neural networks for classification of breast cancer histology images. *Neurocomputing*, 366:46–53, 2019.

[156] Yuchao Zheng, Chen Li, Xiaomin Zhou, Haoyuan Chen, Hao Xu, Yixin Li, Haiqing Zhang, Xiaoyan Li, Hongzan Sun, Xinyu Huang, et al. Application of transfer learning and ensemble learning in image-level classification for breast histopathology. *Intelligent Medicine*, 3(02):115–128, 2023.

[157] Kalpana George, Shameer Faziludeen, Praveen Sankaran, et al. Breast cancer detection from biopsy images using nucleus guided transfer learning and belief based fusion. *Computers in Biology and Medicine*, 124:103954, 2020.

[158] Guangli Li, Chuanxiu Li, Guangting Wu, Guangxin Xu, Ying Zhou, and Hongbin Zhang. Mf-omkt: Model fusion based on online mutual knowledge transfer for breast cancer histopathological image classification. *Artificial Intelligence in Medicine*, 134:102433, 2022.

[159] Yiping Zhou, Can Zhang, and Shaoshuai Gao. Breast cancer classification from histopathological images using resolution adaptive network. *IEEE Access*, 10:35977–35991, 2022.

[160] Shahan Yamin Siddiqui, Amir Haider, Taher M Ghazal, Muhammad Adnan Khan, Iftikhar Naseer, Sagheer Abbas, Muhibur Rahman, Junaid Ahmad Khan, Munir Ahmad, Mohammad Kamrul Hasan, et al. Iomt cloud-based intelligent prediction of breast cancer stages empowered with deep learning. *IEEE Access*, 9:146478–146491, 2021.

[161] Mohiuddin Ahmed and Md Rabiul Islam. A combined feature-vector based multiple instance learning convolutional neural network in breast cancer classification from histopathological images. *Biomedical Signal Processing and Control*, 84:104775, 2023.

[162] Soham Chattopadhyay, Arijit Dey, Pawan Kumar Singh, and Ram Sarkar. Drda-net: Dense residual dual-shuffle attention network for breast cancer classification using histopathological images. *Computers in biology and medicine*, 145:105437, 2022.

[163] Pin Wang, Qi Song, Yongming Li, Shanshan Lv, Jiaxin Wang, Linyu Li, and HeHua Zhang. Cross-task extreme learning machine for breast cancer image classification with deep convolutional features. *Biomedical Signal Processing and Control*, 57:101789, 2020.

[164] Alaa Hussein Abdulaal, Morteza Valizadeh, Mehdi Chehel Amirani, and AFM Shahen Shah. A self-learning deep neural network for classification of breast histopathological images. *Biomedical Signal Processing and Control*, 87:105418, 2024.

[165] Ritesh Maurya, Nageshwar Nath Pandey, Malay Kishore Dutta, and Mohan Karnati. Fccs-net: Breast cancer classification using multi-level fully convolutional-channel and spatial attention-based transfer learning approach. *Biomedical Signal Processing and Control*, 94:106258, 2024.

[166] Weiming Mi, Junjie Li, Yucheng Guo, Xinyu Ren, Zhiyong Liang, Tao Zhang, and Hao Zou. Deep learning-based multi-class classification of breast digital pathology images. *Cancer Management and Research*, pages 4605–4617, 2021.

[167] Salini S Nair and M Subaji. Automated identification of breast cancer type using novel multipath transfer learning and ensemble of classifier. *IEEE Access*, 2024.

[168] Xiaomei Wang, Ijaz Ahmad, Danish Javeed, Syeda Armana Zaidi, Fahad M Alotaibi, Mohamed E Ghoneim, Yousef Ibrahim Daradkeh, Junaid Asghar, and Elsayed Tag Eldin. Intelligent hybrid deep learning model for breast cancer detection. *Electronics*, 11(17):2767, 2022.

[169] Vivek Patel, Vijayshri Chaurasia, Rajesh Mahadeva, and Shashikant P Patole. Garl-net: Graph based adaptive regularized learning deep network for breast cancer classification. *IEEE Access*, 11:9095–9112, 2023.

[170] Soumya Sara Koshy and L Jani Anbarasi. Lmhistnet: Levenberg–marquardt based deep neural network for classification of breast cancer histopathological images. *IEEE Access*, 2024.

[171] Xin Yu Liew, Nazia Hameed, and Jeremie Clos. An investigation of xgboost-based algorithm for breast cancer classification. *Machine Learning with Applications*, 6:100154, 2021.

[172] Jiann-Shu Lee and Wen-Kai Wu. Breast tumor tissue image classification using diu-net. *Sensors*, 22(24):9838, 2022.

[173] Alireza Maleki, Mohammad Raahemi, and Hamid Nasiri. Breast cancer diagnosis from histopathology images using deep neural network and xgboost. *Biomedical Signal Processing and Control*, 86:105152, 2023.

[174] Sreedhar Kollem, Chandrasekhar Sirigiri, and Samineni Peddakrishna. A novel hybrid deep cnn model for breast cancer classification using lipschitz-based image augmentation and recursive feature elimination. *Biomedical Signal Processing and Control*, 95:106406, 2024.

[175] Alper Aksac, Douglas J Demetrick, Tansel Ozyer, and Reda Alhajj. Brecahad: a dataset for breast cancer histopathological annotation and diagnosis. *BMC research notes*, 12:1–3, 2019.

[176] Camelyon 2016. Camelyon 2016: Grand challenge on cancer metastasis detection in lymph nodes. https://camelyon16.grand-challenge.org/, 2016. Accessed: 2024-10-28.

[177] Xueqin Zhang, Chang Liu, Tianren Li, and Yunlan Zhou. The whole slide breast histopathology image detection based on a fused model and heatmaps. *Biomedical Signal Processing and Control*, 82:104532, 2023.

[178] Sepideh Khaliliboroujeni, Xiangjian He, Wenjing Jia, and Saeed Amirgholipour. End-to-end metastasis detection of breast cancer from histopathology whole slide images. *Computerized Medical Imaging and Graphics*, 102:102136, 2022.

[179] Chengyang Gao, Qiule Sun, Wen Zhu, Lizhi Zhang, Jianxin Zhang, Bin Liu, and Junxing Zhang. Transformer based multiple instance learning for wsi breast cancer classification. *Biomedical Signal Processing and Control*, 89:105755, 2024.

[180] Bastiaan S Veeling, Jasper Linmans, Jim Winkens, Taco Cohen, and Max Welling. Rotation equivariant cnns for digital pathology. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*, pages 210–218. Springer, 2018.

[181] Daniel S Luz, Thiago JB Lima, Romuere RV Silva, Deborah MV Magalhães, and Flavio HD Araujo. Automatic detection metastasis in breast histopathological images based on ensemble learning and color adjustment. *Biomedical Signal Processing and Control*, 75:103564, 2022.

[182] Ertan Bütün, Murat Uçan, and Mehmet Kaya. Automatic detection of cancer metastasis in lymph node using deep learning. *Biomedical Signal Processing and Control*, 82:104564, 2023.

[183] Babak E. Bejnordi, Mitko Veta, Paul J. van Diest, Bram van Ginneken, Nico Karssemeijer, and Geert Litjens. Mitos-atypia-14 grand challenge on mitosis detection and nuclear atypia scoring. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*. ICPR, 2014.

[184] M Sreeraj and Jestin Joy. A machine learning based framework for assisting pathologists in grading and counting of breast cancer cells. *ICT Express*, 7(4):440–444, 2021.

[185] Mobeen Ur Rehman, Suhail Akhtar, Muhammad Zakwan, and Muhammad Habib Mahmood. Novel architecture with selected feature vector for effective classification of mitotic and non-mitotic cells in breast cancer histology images. *Biomedical Signal Processing and Control*, 71:103212, 2022.

[186] Guilherme Aresta, Teresa Araújo, Scotty Kwok, Sai Saketh Chennamsetty, Mohammed Safwan, Varghese Alex, Bahram Marami, Marcel Prastawa, Monica Chan, Michael Donovan, et al. Bach: Grand challenge on breast cancer histology images. *Medical image analysis*, 56:122–139, 2019.

[187] Paul Mooney. Breast histopathology images. https://www.kaggle.com/datasets/paultimothymooney/breast-histopathology-images, 2018. Accessed: 13 July 2022.

[188] David Murcia-Gomez, Ignacio Rojas-Valenzuela, and Olga Valenzuela. Impact of image preprocessing methods and deep learning models for classifying histopathological breast cancer images. *Applied Sciences*, 12(22):11375, 2022.

[189] Saman Khalil, Uroosa Nawaz, Zubariah, Zohaib Mushtaq, Saad Arif, Muhammad Zia ur Rehman, Muhammad Farrukh Qureshi, Abdul Malik, Adham Aleid, and Khalid Alhussaini. Enhancing ductal carcinoma classification using transfer learning with 3d u-net models in breast cancer imaging. *Applied Sciences*, 13(7):4255, 2023.

[190] Mamoona Humayun, Muhammad Ibrahim Khalil, Saleh Naif Almuayqil, and Noor Zaman Jhanjhi. Framework for detecting breast cancer risk presence using deep learning. *Electronics*, 12(2):403, 2023.

[191] Michel Gautherie and Charles M Gros. Breast thermography and cancer risk prediction. *Cancer*, 45(1):51–56, 1980.

[192] Manasi B Rakhunde, Shashank Gotarkar, and Sonali G Choudhari. Thermography as a breast cancer screening technique: a review article. *Cureus*, 14(11), 2022.

[193] LF Silva, DCM Saade, GO Sequeiros, AC Silva, AC Paiva, RS Bravo, and Aura Conci. A new database for breast research with infrared image. *Journal of Medical Imaging and Health Informatics*, 4(1):92–100, 2014.

[194] Fayez AlFayez, Mohamed W Abo El-Soud, and Tarek Gaber. Thermogram breast cancer detection: A comparative study of two machine learning techniques. *Applied Sciences*, 10(2):551, 2020.

[195] Samar M. Alqhtani. Breastcnn: A novel layer-based convolutional neural network for breast cancer diagnosis in dmr-thermogram images. *Applied Artificial Intelligence*, 36(1):2067631, 2022.

[196] Areej A Malibari, Marwa Obayya, Abdulbaset Gaddah, Amal S Mehanna, Manar Ahmed Hamza, Mohamed Ibrahim Alsaid, Ishfaq Yaseen, and Amgad Atta Abdelmageed. Artificial hummingbird algorithm with transfer-learning-based mitotic nuclei classification on histopathologic breast cancer images. *Bioengineering*, 10(1):87, 2023.

[197] Herry Suprajitno et al. Investigations on impact of feature normalization techniques for prediction of hydro-climatology data using neural network backpropagation with three layer hidden. *International Journal of Sustainable Development & Planning*, 17(7), 2022.

[198] Nazanin Vafaei, Rita A Ribeiro, and Luis M Camarinha-Matos. Comparison of normalization techniques on data sets with outliers. *International Journal of Decision Support System Technology (IJDSST)*, 14(1):1–17, 2022.

[199] Roseline Oluwaseun Ogundokun, Sanjay Misra, Mychal Douglas, Robertas Damaševičius, and Rytis Maskeliūnas. Medical internet-of-things based breast cancer diagnosis using hyperparameter-optimized neural networks. *Future Internet*, 14(5):153, 2022.

[200] Yaozhong Luo, Zhenkun Lu, Longzhong Liu, and Qinghua Huang. Deep fusion of human-machine knowledge with attention mechanism for breast cancer diagnosis. *Biomedical Signal Processing and Control*, 84:104784, 2023.

[201] Luana Conte, Benedetta Tafuri, Maurizio Portaluri, Alessandro Galiano, Eleonora Maggiulli, and Giorgio De Nunzio. Breast cancer mass detection in dce–mri using deep-learning features followed by discrimination of infiltrative vs. in situ carcinoma through a machine-learning approach. *Applied Sciences*, 10(17):6109, 2020.

[202] Yuvaraja Thangavel, Hitendra Garg, Manjunathan Alagarsamy, and D Pradeep. Revolutionizing breast cancer diagnosis with a comprehensive approach using digital mammogram-based feature extraction and selection for early-stage identification. *Biomedical Signal Processing and Control*, 94:106268, 2024.

[203] Nusrat Ameen Barsha, Aimon Rahman, and MRC Mahdy. Automated detection and grading of invasive ductal carcinoma breast cancer using ensemble of deep learning models. *Computers in Biology and Medicine*, 139:104931, 2021.

[204] Sanae Borrohou, Rachida Fissoune, and Hassan Badir. Data cleaning survey and challenges–improving outlier detection algorithm in machine learning. *Journal of Smart Cities and Society*, 2(3):125–140, 2023.

[205] Mostafa Shanbehzadeh, Hadi Kazemi-Arpanahi, Mohammad Bolbolian Ghalibaf, and Azam Orooji. Performance evaluation of machine learning for breast cancer diagnosis: A case study. *Informatics in Medicine Unlocked*, 31:101009, 2022.

[206] Brindha Senthilkumar, Doris Zodinpuii, Lalawmpuii Pachuau, Saia Chenkual, John Zohmingthanga, Nachimuthu Senthil Kumar, and Lal Hmingliana. Ensemble modelling for early breast cancer prediction from diet and lifestyle. *IFAC-PapersOnLine*, 55(1):429–435, 2022.

[207] Jorge R Vergara and Pablo A Estévez. A review of feature selection methods based on mutual information. *Neural computing and applications*, 24:175–186, 2014.

[208] Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P Trevino, Jiliang Tang, and Huan Liu. Feature selection: A data perspective. *ACM computing surveys (CSUR)*, 50(6):1–45, 2017.

[209] Amir Ahmad and Lipika Dey. A feature selection technique for classificatory analysis. *Pattern Recognition Letters*, 26(1):43–56, 2005.

[210] Zhitao Zhang, Huan Lan, and Shuai Zhao. Analysis of the value of quantitative features in multimodal mri images to construct a radio-omics model for breast cancer diagnosis. *Breast Cancer: Targets and Therapy*, pages 305–318, 2024.

[211] Dongxiao Gu, Kaixiang Su, and Huimin Zhao. A case-based ensemble learning system for explainable breast cancer recurrence prediction. *Artificial Intelligence in Medicine*, 107:101858, 2020.

[212] Haobang Liang, Jiao Li, Hejun Wu, Li Li, Xinrui Zhou, and Xinhua Jiang. Mammographic classification of breast cancer microcalcifications through extreme gradient boosting. *Electronics*, 11(15):2435, 2022.

[213] Karen Simonyan. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[214] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2):301–320, 2005.

[215] Xinyu Liu, Peng Yuan, Ruolin Li, Dejun Zhang, Junda An, Jie Ju, Chenyang Liu, Fuquan Ren, Rui Hou, Yushuang Li, et al. Predicting breast cancer recurrence and metastasis risk by integrating color and texture features of histopathological images and machine learning technologies. *Computers in biology and medicine*, 146:105569, 2022.

[216] R Vijayarajeswari, P Parthasarathy, S Vivekanandan, and A Alavudeen Basha. Classification of mammogram for early detection of breast cancer using svm classifier and hough transform. *Measurement*, 146:800–805, 2019.

[217] Ziyi Liu, Sijie Ni, Chunmei Yang, Weihao Sun, Deqing Huang, Hu Su, Jian Shu, and Na Qin. Axillary lymph node metastasis prediction by contrast-enhanced computed tomography images for breast cancer patients based on deep learning. *Computers in Biology and Medicine*, 136:104715, 2021.

[218] Vidhya Anbalagan and Vanathi Balasubramanian. Hbo-gmrnn: Honey badger optimization based gain modulated recurrent neural network for classification of breast cancer. *Biomedical Signal Processing and Control*, 91:105910, 2024.

[219] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[220] Hossein Talebi and Peyman Milanfar. Learning to resize images for computer vision tasks. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 497–506, 2021.

[221] David Clement, Emmanuel Agu, John Obayemi, Steve Adeshina, and Wole Soboyejo. Breast cancer tumor classification using a bag of deep multi-resolution convolutional features. In *Informatics*, volume 9, page 91. MDPI, 2022.

[222] Rong-Ho Lin, Benjamin Kofi Kujabi, Chun-Ling Chuang, Ching-Shun Lin, and Chun-Jen Chiu. Application of deep learning to construct breast cancer diagnosis model. *Applied Sciences*, 12(4):1957, 2022.

[223] Jyothi Peta and Srinivas Koppu. Explainable soft attentive efficientnet for breast cancer classification in histopathological images. *Biomedical Signal Processing and Control*, 90:105828, 2024.

[224] Adlin Sheeba, P Santhosh Kumar, M Ramamoorthy, and S Sasikala. Microscopic image analysis in breast cancer detection using ensemble deep learning architectures integrated with web of things. *Biomedical Signal Processing and Control*, 79:104048, 2023.

[225] Kushangi Atrey, Bikesh Kumar Singh, Narendra K Bodhey, and Ram Bilas Pachori. Mammography and ultrasound based dual modality classification of breast cancer using a hybrid deep learning approach. *Biomedical Signal Processing and Control*, 86:104919, 2023.

[226] Anas Bilal, Azhar Imran, Xiaowen Liu, Xiling Liu, Zohaib Ahmad, Muhammad Shafiq, Ahmed M El-Sherbeeny, and Haixia Long. Bc-qnet: A quantum-infused elm model for breast cancer diagnosis. *Computers in Biology and Medicine*, 175:108483, 2024.

[227] Jingqi Song, Yuanjie Zheng, Jing Wang, Muhammad Zakir Ullah, Xuecheng Li, Zhenxing Zou, and Guocheng Ding. Multi-feature deep information bottleneck network for breast cancer classification in contrast enhanced spectral mammography. *Pattern Recognition*, 131:108858, 2022.

[228] Prabhpreet Kaur, Gurvinder Singh, and Parminder Kaur. Intellectual detection and validation of automated mammogram breast cancer images by multi-class svm using deep learning classification. *Informatics in Medicine Unlocked*, 16:100151, 2019.

[229] P Esther Jebarani, N Umadevi, Hien Dang, and Marc Pomplun. A novel hybrid k-means and gmm machine learning model for breast cancer detection. *IEEE Access*, 9:146153–146162, 2021.

[230] Naveed Chouhan, Asifullah Khan, Jehan Zeb Shah, Mazhar Hussnain, and Muhammad Waleed Khan. Deep convolutional neural network and emotional learning based breast cancer detection using digital mammography. *Computers in Biology and Medicine*, 132:104318, 2021.

[231] Phillip Chlap, Hang Min, Nym Vandenberg, Jason Dowling, Lois Holloway, and Annette Haworth. A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5):545–563, 2021.

[232] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[233] Steven W Smith. The scientist and engineer's guide to digital signal processing. *California Technical Pub*, 1997.

[234] Yan-Wei Lee, Chiun-Sheng Huang, Chung-Chih Shih, and Ruey-Feng Chang. Axillary lymph node metastasis status prediction of early-stage breast cancer using convolutional neural networks. *Computers in Biology and Medicine*, 130:104206, 2021.

[235] Zaneta Swiderska-Chadaj, Jaime Gallego, Lucia Gonzalez-Lopez, and Gloria Bueno. Detection of ki67 hot-spots of invasive breast cancer based on convolutional neural networks applied to mutual information of h&e and ki67 whole slide images. *Applied Sciences*, 10(21):7761, 2020.

[236] Sameen Aziz, Kashif Munir, Ali Raza, Mubarak S Almutairi, and Shoaib Nawaz. Ivnet: Transfer learning based diagnosis of breast cancer grading using histopathological images of infected cells. *IEEE Access*, 11:127880–127894, 2023.

[237] Deshmukh Pramod Bhausaheb and Kanchan Lata Kashyap. Shuffled shepherd deer hunting optimization based deep neural network for breast cancer classification using breast histopathology images. *Biomedical Signal Processing and Control*, 83:104570, 2023.

[238] Syed Tahir Hussain Rizvi. Real-world applications of machine learning. In *Workshop Proceedings of the 19th International Conference on Intelligent Environments (IE2023)*, pages 98–103. IOS Press, 2023.

[239] Vijayalakshmi Sarraju, Jaya Pal, Supreeti Kamilya, et al. Performance analysis of supervised learning algorithms on different applications. In *CS & IT Conference Proceedings*, volume 12. CS & IT Conference Proceedings, 2022.

[240] Marco Alì, Natascha Claudia D'Amico, Matteo Interlenghi, Marina Maniglio, Deborah Fazzini, Simone Schiaffino, Christian Salvatore, Isabella Castiglioni, and Sergio Papa. A decision support system based on bi-rads and radiomic classifiers to reduce false positive breast calcifications at digital breast tomosynthesis: a preliminary study. *Applied Sciences*, 11(6):2503, 2021.

[241] Yu Li, Chao Huang, Lizhong Ding, Zhongxiao Li, Yijie Pan, and Xin Gao. Deep learning in bioinformatics: Introduction, application, and perspective in the big data era. *Methods*, 166:4–21, 2019.

[242] Abeer Saber, Mohamed Sakr, Osama M Abo-Seida, Arabi Keshk, and Huiling Chen. A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique. *IEEe Access*, 9:71194–71209, 2021.

[243] Yang Li and Shaoying Liu. Adversarial attack and defense in breast cancer deep learning systems. *Bioengineering*, 10(8):973, 2023.

[244] Ella Mahoro and Moulay A Akhloufi. Applying deep learning for breast cancer detection in radiology. *Current Oncology*, 29(11):8767–8793, 2022.

[245] Abdelnour Boukaache, Benhassıne Nasser Edinne, and Djalil Boudjehem. Breast cancer image classification using convolutional neural networks (cnn) models. *International Journal of Informatics and Applied Mathematics*, 6(2):20–34, 2024.

[246] Yan Zhu, Cangzhi Jia, Fuyi Li, and Jiangning Song. Inspector: a lysine succinylation predictor based on edited nearest-neighbor undersampling and adaptive synthetic oversampling. *Analytical biochemistry*, 593:113592, 2020.

[247] Alex X Wang, Stefanka S Chukova, and Binh P Nguyen. Synthetic minority oversampling using edited displacement-based k-nearest neighbors. *Applied Soft Computing*, 148:110895, 2023.

[248] Ashhadul Islam, Samir Brahim Belhaouari, Atiq Ur Rehman, and Halima Bensmail. Knnor: An oversampling technique for imbalanced datasets. *Applied soft computing*, 115:108288, 2022.

[249] Samir Brahim Belhaouari, Ashhadul Islam, Khelil Kassoul, Ala Al-Fuqaha, and Abdesselam Bouzerdoum. Oversampling techniques for imbalanced data in regression. *Expert Systems with Applications*, 252:124118, 2024.

[250] İsmet Abacı and Kazım Yıldız. Smote vs. knnor: An evaluation of oversampling techniques in machine learning. *Gümüşhane Üniversitesi Fen Bilimleri Dergisi*, 13(3):767–779, 2023.

[251] Bo Sun, Qian Zhou, Zhijun Wang, Peng Lan, Yunsheng Song, Shaomin Mu, Aifeng Li, Haiyan Chen, and Peng Liu. Radial-based undersampling approach with adaptive undersampling ratio determination. *Neurocomputing*, 553:126544, 2023.

[252] Dalwinder Singh and Birmohan Singh. Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97:105524, 2020.

[253] Sajjad Nematzadeh, Farzad Kiani, Mahsa Torkamanian-Afshar, and Nizamettin Aydin. Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: A bioinformatics study on biomedical and biological cases. *Computational biology and chemistry*, 97:107619, 2022.

[254] Farbod Farhangi. Investigating the role of data preprocessing, hyperparameters tuning, and type of machine learning algorithm in the improvement of drowsy eeg signal modeling. *Intelligent Systems with Applications*, 15:200100, 2022.

[255] Jiawei Han, Jian Pei, and Hanghang Tong. *Data mining: concepts and techniques*. Morgan kaufmann, 2022.

[256] Muhammad Ali Qureshi, Azeddine Beghdadi, and Mohamed Deriche. Towards the design of a consistent image contrast enhancement evaluation measure. *Signal Processing: Image Communication*, 58:212–227, 2017.

[257] Kui Liu, Guixia Kang, Ningbo Zhang, and Beibei Hou. Breast cancer classification based on fully-connected layer first convolutional neural networks. *IEEE Access*, 6:23722–23732, 2018.

[258] Usman Naseem, Junaid Rashid, Liaqat Ali, Jungeun Kim, Qazi Emad Ul Haq, Mazhar Javed Awan, and Muhammad Imran. An automatic detection of breast cancer diagnosis and prognosis based on machine learning using ensemble of classifiers. *Ieee Access*, 10:78242–78252, 2022.

[259] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[260] Nahid Ferdous Aurna, Mohammad Abu Yousuf, Kazi Abu Taher, AKM Azad, and Mohammad Ali Moni. A classification of mri brain tumor based on two stage feature level ensemble of deep cnn models. *Computers in biology and medicine*, 146:105539, 2022.

[261] Mohammad Amin Morid, Alireza Borjali, and Guilherme Del Fiol. A scoping review of transfer learning research on medical image analysis using imagenet. *Computers in biology and medicine*, 128:104115, 2021.

[262] Samriddha Majumdar, Payel Pramanik, and Ram Sarkar. Gamma function based ensemble of cnn models for breast cancer detection in histopathology images. *Expert Systems with Applications*, 213:119022, 2023.

[263] Ahmet Kursad Poyraz, Sengul Dogan, Erhan Akbal, and Turker Tuncer. Automated brain disease classification using exemplar deep features. *Biomedical Signal Processing and Control*, 73:103448, 2022.

[264] M Emin Sahin. Deep learning-based approach for detecting covid-19 in chest x-rays. *Biomedical Signal Processing and Control*, 78:103977, 2022.

[265] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[266] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[267] Kashif Shaheed, Aihua Mao, Imran Qureshi, Munish Kumar, Sumaira Hussain, Inam Ullah, and Xingming Zhang. Ds-cnn: A pre-trained xception model based on depth-wise separable convolutional neural network for finger vein recognition. *Expert Systems with Applications*, 191:116288, 2022.

[268] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.

[269] Oliver Kramer and Oliver Kramer. K-nearest neighbors. *Dimensionality reduction with unsupervised nearest neighbors*, pages 13–23, 2013.

[270] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[271] Xiaofeng Qi, Fasheng Yi, Lei Zhang, Yao Chen, Yong Pi, Yuanyuan Chen, Jixiang Guo, Jianyong Wang, Quan Guo, Jilan Li, et al. Computer-aided diagnosis of breast cancer in ultrasonography images by deep learning. *Neurocomputing*, 472:152–165, 2022.

[272] Michal Byra. Breast mass classification with transfer learning based on scaling of deep representations. *Biomedical Signal Processing and Control*, 69:102828, 2021.

[273] Ahmed Ali and Maria Khan. Transfer learning using pre-trained cnns for breast cancer image analysis. *Medical Image Analysis*, 67:78–90, 2023.

[274] Priya Radhakrishnan and Rohit Kumar. Transfer learning-based cnn models for breast cancer prediction: A comparative study. *IEEE Transactions on Artificial Intelligence*, 12(5):456–467, 2023.

[275] Pablo Guerra and Sofia Martinez. Transfer learning approaches in breast cancer diagnosis using ultrasound images. *Expert Systems with Applications*, 72:234–246, 2023.

[276] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[277] John Kurama and L. Davis. Review of machine learning approaches for breast cancer diagnosis. *Machine Learning and Applications*, 32:123–135, 2020.

[278] Mikhail Ashurov and Ayesha Zainab. Concatenation of multi-level features for robust breast cancer classification. *Medical Image Computing and Applications*, 50:456–467, 2024.

[279] A. Saad and F. Ahmed. Covid-19 detection using convolutional neural networks: Lessons for breast cancer imaging. *Journal of Medical Informatics*, 30:1–15, 2022.

[280] Selim Cengil and Duygu Turan. Effect of feature selection on breast cancer diagnosis using machine learning models. *Computers in Biology and Medicine*, 150:106345, 2022.

[281] Yasmine El Kassas and Others. Concatenation-based deep learning for breast cancer image analysis. *Journal of AI in Medicine*, 12(2):345–356, 2021.

[282] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

[283] Marta Coelho and Pedro Almeida. Lasso regression for feature selection in breast cancer datasets. *International Journal of Machine Learning Applications*, 15(3):67–78, 2020.

[284] Wei Chen, Li Zhang, et al. Multi-modal deep learning framework for breast cancer diagnosis. *IEEE Transactions on Medical Imaging*, 43:45–57, 2024.

[285] Pin Wang, Jiaxin Wang, Yongming Li, Pufei Li, Linyu Li, and Mingfeng Jiang. Automatic classification of breast cancer histopathological images based on deep feature fusion and enhanced routing. *Biomedical Signal Processing and Control*, 65:102341, 2021.