

Name: Ahmad Moneer Mzayek

Course Title: Research Methodology in Computer Science

Course Code: AI800

Lecturer: Dr. Mohammed Anbar

Hand-in Method: Online Submission via E-Learning.

Assignment Report: Breast cancer detection and diagnosis using artificial intelligence - 3

Contents

Chapter 1: Abstract	3
Chapter 2: Introduction.....	3
Chapter 3: Literature Review	5
Chapter 4 : Research Methodology.....	10
Chapter 5: References	13

Chapter 1: Abstract

Breast cancer remains a significant global health challenge as one of the most frequent and deadly cancers among females. Early detection with high specificity is an essential component to improve prognosis and decrease mortality. Despite improvements in traditional diagnostics using mammography and MRI, such exams are highly dependent on the knowledge of radiologists, making them prone to variability in human interpretation with resulting false positives and negatives. AI, more specifically deep learning, has the potential to diagnosis of breast cancer by automation, enhancement of accuracy, and provision of personalized information. However, most current AI models face serious challenges: low interpretability, a high false-positive rate, and unable to generalize across diverse populations. This research, therefore, proposes a hybrid AI framework that incorporates DL with radionics-based approaches, utilizing multimodal imaging and interpretability tools for improved diagnostic accuracy and usability. The current study aims to develop a robust AI system that not only increases clinical decision-making but also gains trust among healthcare providers for widespread confirmed accuracy, interpretability, and generalizability for breast cancer care globally by addressing dataset variability, ethical concerns, and operational challenges.

Chapter 2: Introduction

- **Introduction**

Breast cancer is one of the major causes of cancer deaths worldwide. It has been estimated that millions of new cases are diagnosed each year. The earlier the diagnosis, the better the prognosis and treatment. While usual mammography, ultrasound, and MRI have been important for the early detection of cancers, there are large limitations in all these modalities due to dependence on radiologists' interpretations, which are variable and prone to human error. The consequences are false positives, leading to unnecessary stress and interruption, while false negatives delay important treatments. AI, and more especially DL, has been emerging as a transformative technology in medical imaging. It holds promise for achieving human performance and even passing it in recognizing complex patterns and abnormality. The subset of DL, namely CNN, has been applied broadly in diagnostics related to breast cancer and show striking capabilities with regard to tumor detection and classification. otherwise these advances, a number of challenges remain, including a high false-positive rate, an absence of interpretability, and lack of generalizability of AI algorithms to routine clinical practice. This study investigates the creation of a hybrid AI approach in order to overcome those obstacles and to improve diagnostic performance for clinicians and patient outcomes.

- **Background**

Medical imaging, especially in the diagnosis of breast cancer, has changed with the introduction of AI. Breast cancer is the most common malignancy among women and the second leading cause of tumor deaths worldwide. Traditional detection techniques include mammography and MRI, which are still important but highly dependent on radiologists' skills and often result in variability in diagnostic performance. Most human control leads to false positives, which create unnecessary stress in patients, and false negatives delay lifesaving interventions.

The application of AI that can process large amounts of data and reveal patterns invisible to the human eye promises a solution. For example, CNNs have showed unparalleled performance in image classification. but, several limitations continue with most existing AI models. They are most often trained on homogeneous datasets, compromising their generalizability in diverse populations. The black-box nature of DL models also limits their interpretability, reducing trust amongst clinicians. This study investigated the integration of DL and radiomics-based approaches that aim at creating an AI framework that is accurate, interpretable, and generalizable across various clinical settings.

- **Problem Statement**

While much progress has been achieved in medical imaging, the detection and diagnosis of breast cancer remain at the frontiers of challenge. The usual imaging modalities of mammography, ultrasound, and MRI are extremely dependent on subjective interpretations by radiologists and will lead to human error and variability in diagnostic results. This might result in unnecessary biopsies and psychological trauma due to false-positive diagnoses, while on the other hand, This may lead to unnecessary biopsies and psychological trauma due to false-positive diagnoses, while false-negative diagnoses may result in delayed treatment. Artificial intelligence has the possibility to change the face of diagnostics in breast cancer by automating image interpretation and improving diagnostic accuracy.

However, the current AI models have significant limitations. Most of them are developed using homogeneous datasets, which limits their generalizability across diverse populations and imaging systems. High false-positive rates are an unnecessary interruption and a source of stress for the patients. The opaque processes of the decisions made by DL models further undermine clinicians' trust in integrating these tools into workflows. Radiomics-based approaches often lack the predictive power for effective early detection, whereas being interpretable. The said research work, therefore, will try to address identified challenges by developing an AI framework that integrates accuracy and interpretability with generalizability to revolutionize diagnostics of breast cancer.

- **Research objective**

This research aims to propose the next-generation AI framework that overcomes the existing gaps related to accuracy, interpretability, and generalizability issues of breast cancer diagnosis and detection. The following is a summary of the main specific objectives of this study:

planning a hybrid AI model by integrating the power of CNNs with that of radiomics-based approaches to leverage their respective strengths for extended diagnostic accuracy.

The integration of multi-imaging data, such as mammography, MRI, and DCE-MRI, into the AI framework to improve detection performance and review of the impact of multicenter datasets

The development of an interpretability framework that shall provide clinicians with clear insights into AI decision-making is required for building trust and usability in clinical settings. It will be necessary to develop actionable guidelines on how to integrate the AI framework into existing workflows, considering regulatory, ethical, and operational issues.

In light of these objectives, the research seeks to build an AI system that will revolutionize the diagnosis of breast cancer, improve outcomes for patients, and make healthcare systems more effective around the world.

- **Research Scope**

The work presented herein focuses on applying AI techniques for the better detection of breast cancer by considering diverse datasets and multimodal imaging technologies. It covers some technical challenges-interpretability and generalization-along with practical ones, like clinical usability, ethical issues, and regulatory compliance.

Chapter 3: Literature Review

- **Study papers**

Paper 1 : Deep-learning based breast cancer detection for cross-staining histopathology images

This research seeks to overcome the challenges posed by cross-staining variation in histopathology images for the detection of breast cancer. It thus combines H&E staining with fluorescent staining to construct a deep learning framework that can recognize tumors regardless of the staining method used. The primary objective is to normalize color variations and extract shared features for consistent analysis using computational pathology models.

It involves three different preprocessing workflows comprising the CD and ACS algorithms in dealing with staining discrepancies. They then used MobileNetV2 with a U-Net decoding structure for image segmentation. These models were trained and then tested on datasets of images containing 30 H&E-stained slides and 30 fluorescently stained slides annotated by biologists and validated by pathologists. Performance was evaluated using metrics of accuracy, sensitivity, and specificity within the staining modalities.

Paper 2: Recent Advancements in Machine Learning and Deep Learning-based Breast Cancer Detection Using Mammograms

The paper reviews the recent developments in ML and DL-based methodologies for the detection of breast cancer using mammograms. Emphasis is placed on the development of appropriate diagnostic and classification methodologies that will be useful in early cancer detection, with a focus on the role of artificial intelligence to help overcome challenges in mammography.

Paper 3: Machine Learning, IoT, and 5G Technologies for Breast Cancer Studies: A Review

The paper reviews integrating Machine learning , Internet of things , and 5G technologies in the diagnosis, treatment, and management of BC patients. It considers recent developments in the field of computer-aided diagnostic systems and discusses how emerging IoT and 5G improve real-time monitoring, sharing, and management of patients.

Paper 4 : AI-Based Hierarchical Approach for Optimizing Breast Cancer Detection Using MammoWave Device

The study It describes and assesses an AI-driven hierarchical approach to optimize the detection of breast cancer by means of the MammoWave device, a microwave imaging system. The idea is to transcend some limitations of traditional mammography, especially for younger women and those whose testing is limited due to radiation concerns, by using a safe and innovative technology.

- **Critical Analysis**

The 4 papers collectively contribute to the newest methodologies for breast cancer detection with AI, IoT, 5G, and advanced imaging. Paper 1 and Paper 3 develop deep learning (DL) that achieve the best diagnostic performance through image segmentation, feature extraction, and classification. Paper 2 broadens the technological horizon by integrating IoT and 5G for real-time monitoring and seamless data sharing, while Paper 4 presents a pioneering cross-staining recognition algorithm to bridge the gap between traditional and fluorescent histopathology imaging. At the same time, all these papers share one major weakness related to data availability and diversity, hence the limitation of generalizability on wider populations. Papers 3 and 4 present computationally intensive methods, while Paper 2 does not have experiments to prove their theoretical conjectures. Further, ethical considerations, ensuring data security, and meeting infrastructural needs are lacunae in all the studies performed, therefore pointing to the yawning chasm between research originality and clinical translation.

- **Connections**

The 4 papers reinforce the potential of AI to improve the detection of breast cancer, all having a common agenda of using technology to improve the diagnostic performance for early treatment. Papers 1, 3, and 4 meet on the power of DL to face challenges in image analysis, while Paper 2 integrating IoT and 5G for broader healthcare accessibility. Shared recognition of dataset limitations ties the papers together. All studies seek to bridge technological advancements with clinical applications, reflecting a common goal of optimizing patient outcomes through early and accurate diagnosis.

- **Conflicts**

While sharing a related purpose, the papers diverge in approach and priorities: Papers 1, 3, and 4 dwell on AI and imaging advances for enhanced diagnostic precision, whereas Paper 2 focuses on IoT and 5G for real-world, scalable healthcare solutions. This difference in focus highlights a possible trade-off between high-performance models in controlled settings and accessibility-driven technologies proposed in Paper 2.. Furthermore, the focus of Paper 4 on resolving inconsistencies in staining underlines challenges in imaging-specific workflows-a direct contradiction to the broader technological integration suggested by Paper 2. Variations thus suggest different ways in which future research might take place and reflect the multidimensionality of the issues to be overcome in the arena of breast cancer detection.

- **Creation of a table summarizing the reviewed papers**

Study	Year	RDk	Key Findings/Results	Methodology	Limitations
Deep-learning based breast cancer detection for cross-staining histopathology images	2024	Article	This work developed a deep-learning framework to handle cross-staining variations in histopathology images, using MobileNetV2 as the encoder and U-Net for segmentation. It has considered three workflows: RGB (raw image data), CD (color deconvolution), and CDACS (that combines CD with adaptive color segmentation to enhance feature extraction). The model was trained on 30 H&E-stained slides and 30 fluorescent-stained slides while the performance was validated using accuracy, sensitivity, and specificity metrics across both staining methods.	A deep-learning framework was developed using MobileNetV2 as an encoder and U-Net segmentation to address cross-staining variations in histopathology images. Three workflows were followed: RGB, raw image data; CD, color deconvolution; and CDACS, a combination of color deconvolution with adaptive color segmentation for better feature extraction. The model is trained on 30 H&E-stained slides and 30 fluorescent-stained slides; its performance is validated on accuracy, sensitivity, and specificity metrics across both staining methods.	Small Dataset: Limited to 30 slides for each staining method, reducing generalizability. Focus on Nuclei: Relied on features in the nuclei, losing tumor heterogeneity. Fluorescent Data with Lower Performance: The accuracy, for fluorescent-stained images, was 80.5%. Limited Real-Life Validation: Retrospective study without prospectively performed clinical testing. Generalization Issues: Variability in staining protocols and imaging equipment was not considered.
Recent Advancements in Machine Learning and Deep Learning-based Breast Cancer Detection Using Mammograms	2024	Article	DL-based methods, Among them, especially CNNs and transfer learning algorithms outperform classical machine learning techniques in terms of classification accuracy and robustness. Transfer learning works very effectively on small datasets, thus making this approach preferable over classical CNNs. Most existing works focus on binary classification (malignant-benign or normal-abnormal) rather than three-class classification (normal, benign, malignant). Enhanced segmentation and feature extraction have significantly improved the detection accuracy while hybrid and	Framework and Categorization: The paper categorizes various breast cancer detection methods, including classical ML and DL techniques. Techniques Reviewed: Image Pre-processing: Techniques to enhance image quality and suppress noise. Segmentation: Methods such as U-Net, YOLO, and hybrid approaches for identifying regions of interest. Feature Extraction and Reduction: Analysis of geometrical, morphological, and texture-based features, including DL-driven features like CNN-based architectures. Classification: Classical ML models (SVM, KNN) and DL	Publicly available datasets are limited in size, which restricts the performance of DL models. Publicly available datasets are limited in size, which restricts the performance of DL models. Limited Classifications: Most studies fail to classify mammograms into three classes, limiting diagnostic granularity. Ground Truth Challenges: The ground truth data for segmentation and classification is rare; therefore, model validation is inconsistent. Variability in Data: Irregularities in breast

			ensemble models bear great potential.	models (ResNet, VGGNet, and transfer learning frameworks).	mass shape, size, and texture make detection and classification more challenging.
Machine Learning, IoT, and 5G Technologies for Breast Cancer Studies: A Review	2024	Article	Indeed, DL outperforms other conventional ML models, especially with big datasets, to result in high accuracy of classification and prognosis of BC. IoT devices can improve personalized care, real-time monitoring, and post-treatment recovery by integrating sensors, wearables, and smart prosthetics. 5G technologies support faster image transfers, remote diagnostics, and access to healthcare in underserved communities. While ML-based CAD systems show promise, their application is hindered by limited data availability and inconsistent performance across diverse datasets.	Machine Learning and Deep Learning Approaches: A detailed review of ML and deep learning (DL) classifiers for BC detection, including SVM, KNN, CNN, and transfer learning models. Imaging modality analysis, including mammograms, MRI, and histopathological images. IoT Applications Wearable devices and sensors for real-time patient monitoring, enabling personalized patient care. Smarter intake and transport of data with IoT-enabled systems. 5G Integration High-speed data transfer and remote diagnostics. Telemedicine and collaborative research on low-latency networks.	The available medical datasets limit model validation and generalization. Annotation Complexity: The manual annotation of imaging data is laborious and prone to errors. Algorithmic Gaps: Many ML models require improvements in accuracy, robustness, and consistency to replace human intervention in clinical settings. Data Privacy Concerns: IoT and 5G use raises significant privacy and security challenges regarding sensitive patient

AI-Based Hierarchical Approach for Optimizing Breast Cancer Detection Using MammoWave Device	2023	Article	The hierarchical approach showed an overall balanced performance of about 70% in sensitivity, specificity, and accuracy. This non-invasive microwave imaging from the MammoWave device, together with the hierarchical model, is a promising alternative to traditional mammography. Sub-band 8, with frequencies from 8 to 9 GHz, turned out to be the most important sub-band for tumor detection. Naive Bayes turned out to be one of the most consistent classifiers by balanced performance, combined with effective feature extraction (SF_M, SF_R) and Normalization Techniques (Min-Max, Robust	Used a dataset of 1024 breast samples from two clinical trials in European hospitals. Hierarchical approach with two main steps: First, unsupervised clustering to classify data. Then, binary classification inside the clusters to identify either healthy or non-healthy cases. Optimization of feature extraction methods, normalization techniques, and classifiers for each step. Feature Normalization and Extraction: The features extracted by different techniques are SF, PCA, tSNE, and IMAP. The pre-processing techniques used for normalization to make the data uniform include Min-Max, Z-score, and Robust scaling.	While diverse, the dataset is still rather small, a limitation when generalizing with the model. Feature and Model Dependence: There may be further dependencies on the selected sub-bands and features, making generalization difficult for larger data sets. Optimization Complexity: Hierarchical and optimization processes are computationally expensive and require extensive fine-tuning.
--	------	---------	--	--	---

Chapter 4: Research Methodology

- **Methodology**

The proposed research adopts a hybrid methodology that combines qualitative and quantitative techniques:

1. **Data Collection:** Multi-center datasets encompassing mammography, MRI, and DCE-MRI.
2. **Model Development:** Designing a hybrid AI model integrating CNNs with radiomics-based approaches.
3. **Validation:** Evaluating the model using metrics such as sensitivity, specificity, and AUC on diverse datasets.
4. **Interpretability Framework:** Developing tools to visualize and explain AI decisions to clinicians.

- **Expected Outcomes**

The research will contribute to the field of diagnostics in breast cancer through the development of an integrated AI framework that incorporates the latest DL techniques and radiomics-based approaches. This is expected to result in the following:

- **Improved Diagnostic Performance:** The proposed hybrid AI framework is very likely to outperform existing systems in diagnostic performance by putting the multimodal imaging data of mammography, MRI, and DCE-MRI to effective use. The model aims at minimizing false positives and false negatives so that malignancies could be found in a far more reliable manner.
- **Improved Generalizability:** It is expected that the generalization of the proposed framework should be very high as multi-center datasets with diverse demographics, different imaging devices, and varied clinical settings will be used in the training. It will be trained on multi-center datasets with diverse demographics, imaging devices, and clinical settings. It will therefore have wide applicability for a variety of patient populations and clinical settings.
- **Improved Interpretability:** Integrating interpretability with the model will provide clinicians with insight into AI decision-making processes and, in turn, build their confidence in integrating this into clinical workflows. These are foreseen to include visualization tools, such as heatmaps and feature visualization, which may make AI predictions more interpretable by practitioners.
- **Clinical Integration Guidelines:** The study will deliver key recommendations on how best the AI framework should be ethically, operationally, and regulatorily integrated into the present clinical practices. Besides this, it shall provide details on data privacy challenges, security concerns, and how to maintain standards operating in health care..
- **Resource Optimization:** By reducing the dependency on radiologists for routine screenings and minimizing unnecessary follow-ups caused by false positives, the framework is expected to optimize healthcare resources, reduce costs, and improve the efficiency of breast cancer diagnostics.
- **Improved Patient Outcomes:** The framework will enhance treatment timelines and survival rates through early detection and the avoidance of diagnostic inaccuracies that are psychologically burdensome.

- **Gantt Chart**

The Gantt chart below illustrates the timeline for the research phases:

Task	11-1	11-15	12-1	12-5	12-12
Literature Review	▼				
Data Collection	▼				
Model Development		▼			
Validation and Testing			▼		
Interpretability Framework				▼	
Final Report and Submission					▼

Chapter 5: References

- 1- <https://www.sciencedirect.com/science/article/pii/S1746809424012011>

By: Mehran Taghipour-Gorjilaie, Navid Ghavami, Lorenzo Papini , Mario Badia ,Arianna Fracassini , Alessandra Bigotti , Gianmarco Palomba ,Daniel ´ Alvarez S´anchez-Bayuela , Cristina Romero Castellano, Riccardo Loretoni,Massimo Calabrese, Alberto Stefano Tagliafico , Mohammad Ghavami , Gianluigi Tiberi.

- 2- <https://www.sciencedirect.com/science/article/pii/S240584402300378X>

By: Pei-Wen Huang, Hsu Ouyang, Bang-Yi Hsu, Yu-Ruei Chang, Yu-Chieh Lin, Yung-An Chen, Yu-Han Hsieh, Chien-Chung Fu, Chien-Feng Li, Ching-Hung Lin, Yen-Yin Lin, Margaret Dah-Tsyr Chang, Tun-Wen Pai.

- 3- <https://www.sciencedirect.com/science/article/pii/S1110016824000541>

By: Havva Elif Saroğlu, Ibraheem Shayea, Bilal Saoud, Marwan Hadri Azmi, Ayman A. El-Saleh, Sawsan Ali Saad, Mohammad Alnakhli.

- 4- <https://www.sciencedirect.com/science/article/pii/S1120179723001655>

by: Adyasha Sahu, Pradeep Kumar Das, Sukadev Meher.

- 5- <https://www.sciencedirect.com/science/article/pii/S2950363924000048>

By Oladosu Oyebisi Oladimejia,b, Ian Mcloughlin, Saritha Unnikrishnan

- 6- <https://www.sciencedirect.com/science/article/pii/S1076633224007748>

By Lukas Hirsch ^a, Yu Huang ^a, Hernan A. Makse ^a, Danny F. Martinez ^b, Mary Hughes ^b, Sarah Eskreis-Winkler ^b, Katja Pinker ^b, Elizabeth A. Morris ^{b c}, Lucas C. Parra ^a, Elizabeth J. Sutton ^b

- 7- https://www.researchgate.net/publication/341556607_Detection_and_Diagnosis_of_Breast_Cancer_Using_Artificial_Intelligence_Based_Assessment_of_Maximum_Intensity_Projection_Dynamic_Contrast-Enhanced_Magnetic_Resonance_Images

By Mio Adachi 1, Tomoyuki Fujioka 2,* , Mio Mori 2 , Kazunori Kubota 2,3 , Yuka Kikuchi 2,Wu Xiaotong 2, Jun Oyama 2, Koichiro Kimura 2, Goshi Oda 1, Tsuyoshi Nakagawa 1,HiroYuki Uetake 1 and Ukihide Tateishi

8- <https://www.sciencedirect.com/science/article/pii/S2667325823001814>

By Qing Lin , Wei-Min Tan , Jing-Yu Ge , Yan Huang , Qin Xiao ,Ying-Ying Xu , Yi-Ting Jin , Zhi-Ming Shao , Ya-Jia Gu , Bo Yan ,Ke-Da Yu

9- <https://www.sciencedirect.com/science/article/pii/S0720048X24001736>

By Oliver Díaz a,b, Alejandro Rodríguez-Ruíz c, Ioannis Sechopoulos

10- <https://www.sciencedirect.com/science/article/pii/S2949916X24000732>

By Meredith A. Jones^{1*}, Warid Islam², Rozwat Faiz²,Xuxin Chen² and Bin Zheng²