

Centralized Multi Modal Deep Learning for Breast Cancer Diagnosis: A Physics Aware Approach

Haneen Qusay Mawlood¹, Ahmed ALkarawi², Roa'a Alı Abdullah Mohammedqasem³

^{1,2,3} Department of Computer Engineering, Istanbul Aydin University, Istanbul, Turkey

ABSTRACT: Breast cancer is still a major cause of cancer death in women around the world, so it needs a precise diagnosis that often goes beyond one imaging method. While deep learning has shown great promise in automated diagnosis, current research often suffers from the separation of imaging modalities. This study offers a comprehensive, physics-aware deep learning framework that combines Ultrasound, MRI, and Mammography. We present custom preprocessing pipelines specific to each modality to tackle the different physical degradation models associated with each type of imaging. For MRI, we adopt a 2D slice-based methodology involving Key Slice Extraction to find the most informative tumor cross-section followed by N4 Bias Field Correction and Otsu's thresholding for intensity normalization. This allows the effective use of 2D Convolutional Neural Networks without incurring the computational cost associated with 3D processing. Using transfer learning from ResNet50 (Ultrasound/Mammography) and DenseNet121 (MRI), our centralized models reached state-of-the-art accuracies of 92.50%, 90.63%, and 92.00%, proving that a centralized multi-modal approach can be effective in enhancing diagnostic precision.

KEYWORDS: Breast Cancer, Deep Learning, Multi-Modality Imaging, 2D Slice Extraction, ResNet50, DenseNet121.

I. INTRODUCTION

1.1 Background and Motivation

Breast cancer accounts for approximately 25% of all new cancer cases in women, making it the most prevalent malignancy among the female population worldwide [19]. The complexity of breast cancer, characterized by its heterogeneous nature, requires a multi-faceted diagnostic approach. Contemporary diagnostic workflows are inherently multi modal, routinely synthesizing information from X ray Mammography (for detecting microcalcifications), Ultrasound (for characterizing tissue density and lesion margins), and Magnetic Resonance Imaging (MRI) (for assessing vascularity and tumor extent) [4]. Despite the availability of these modalities, the manual interpretation of such complex data streams is highly subjective and prone to inter observer variability, which can lead to delayed diagnosis or unnecessary biopsies.

The advent of deep learning (DL) has revolutionized medical imaging by providing automated, high precision tools for lesion detection and classification. However, the clinical deployment of these models faces a significant hurdle: each imaging modality possesses unique physical characteristics and noise profiles—such as speckle noise in ultrasound or intensity inhomogeneity in MRI—that generic DL models often fail to account for [8]. Effective integration of these diverse modalities within a robust deep learning framework is crucial for maximizing diagnostic accuracy.

1.2 Research Gap

A critical gap exists in the current literature regarding the efficient processing of volumetric MRI data. Most existing studies utilize computationally expensive 3D CNNs, which often struggle with convergence on relatively small medical datasets due to the "curse of dimensionality" [10]. Conversely, simplistic 2D approaches that treat MRI slices as independent images often fail to capture the volumetric context of the tumor. Furthermore, while multi modal fusion has shown promise, many frameworks rely on simple feature concatenation without considering the physical principles underlying each modality. There is a need for a comprehensive, physics aware multi modal deep learning framework that efficiently integrates diverse imaging data for breast cancer diagnosis [2].

1.3 Contributions

This study addresses these gaps through the following contributions:

1 **Physics Aware Preprocessing:** We developed tailored pipelines for Ultrasound (Speckle reduction), MRI (Bias field correction on 2D slices), and Mammography (Contrast enhancement), ensuring that the input data respects the physical constraints of the imaging process [8].

2 **Intelligent 2D MRI Methodology:** We validated a robust approach for converting 3D MRI volumes into high quality 2D inputs through Key Slice Extraction, enabling the

use of efficient, pre trained 2D architectures like DenseNet121 [10].

3 Centralized Multi Modal Deep Learning Framework: We propose and validate a centralized deep learning framework that effectively integrates features from Ultrasound, MRI, and Mammography for enhanced breast cancer diagnosis.

II. LITERATURE REVIEW

A. Deep Learning in Single Modality Breast Imaging

The evolution of deep learning in breast cancer diagnosis has seen a transition from basic convolutional architectures to complex, multi scale networks. In Ultrasound imaging, Alom et al. (2025) achieved an accuracy of 89.87% by utilizing an explainable AI driven approach, though their performance was limited by the use of generic preprocessing techniques that did not fully address speckle noise [1]. Similarly, in Mammography, Khan (2021) established a baseline of 88.00% using digital breast tomosynthesis, highlighting the challenges of detecting subtle microcalcifications in dense breast tissue [3]. MRI analysis has traditionally relied on 3D CNNs, but Abdullah and Nuzla (2025) reported that intensity inhomogeneity remains a significant barrier, even with advanced volumetric models, leading to a reported accuracy of 87.50% [2]. Recent advancements have also explored knowledge augmented deep learning to improve diagnostic accuracy [17]. For instance, Guo et al. (2024) demonstrated that augmenting CNNs with symbolic knowledge derived from clinical guidelines can significantly enhance the model's ability to interpret complex mammographic patterns [17].

B. Multi Modal Fusion Strategies

Multi modal fusion aims to integrate complementary information from different sources to improve diagnostic robustness. Li et al. (2025) categorized these strategies into three primary types: feature level, decision level, and hybrid fusion [4].

Feature Level Fusion: This approach involves concatenating or merging features extracted from different modalities at an early stage, typically before the final classification layer. The advantage lies in allowing the model to learn complex cross modal correlations directly from the combined feature space. However, this method often leads to high dimensional feature vectors, increasing computational complexity and the risk of overfitting, especially with limited datasets [20]. Examples include early fusion techniques where raw data or low level features from different modalities are combined into a single representation [28].

Decision Level Fusion: In contrast, decision level fusion operates at a later stage, where independent models are trained for each modality. Their individual outputs (e.g., class probabilities or diagnostic scores) are then combined using various aggregation techniques such as majority voting, weighted averaging, or more sophisticated machine learning classifiers [22]. This strategy offers greater robustness to

missing modalities and allows for the use of specialized models for each data type. However, it may miss subtle inter modal interactions that could be captured at an earlier fusion stage [23].

Hybrid Fusion: Hybrid fusion strategies combine elements of both feature level and decision level fusion, offering a more flexible and comprehensive approach to integrating multi modal data. These methods often involve multiple fusion points throughout the network architecture, allowing for the capture of both local and global dependencies across modalities [4]. Recent work has demonstrated that integrating ultrasound and mammography using a hybrid residual network can significantly boost performance over single modality baselines by leveraging the strengths of both early and late fusion [25]. Furthermore, advanced hybrid fusion networks incorporate attention mechanisms to dynamically weigh the importance of different modalities or features, leading to improved diagnostic accuracy [27].

C. Physics Aware and Knowledge Augmented Learning

Recent trends have shifted towards incorporating domain knowledge and physical principles directly into deep learning frameworks to enhance their performance and interpretability. Physics informed machine learning (PIML) integrates fundamental physical principles—such as the wave equation in ultrasound or magnetic field physics in MRI directly into the neural network's architecture or loss function [8]. This approach helps regularize the learning process, improves the model's interpretability, and enhances generalization capabilities, especially in scenarios where labeled data is scarce or expensive to acquire [9]. For instance, physics guided conditional diffusion networks have been proposed for image reconstruction and enhancement, demonstrating superior performance by embedding physical constraints into the generative process [9]. Furthermore, Knowledge Augmented Deep Learning (KADL) leverages clinical guidelines, expert knowledge, and radiomic features to guide the learning process. Guo et al. (2024) showed that augmenting CNNs with symbolic knowledge derived from clinical guidelines can significantly enhance the model's ability to interpret complex mammographic patterns and improve diagnostic accuracy [17]. This integration of human expertise with data driven models helps to overcome the black box nature of deep learning, making the models more trustworthy and clinically actionable.

III. METHODOLOGY

A. Proposed Centralized Multi Modal Framework

Our proposed framework, illustrated in Figure 1, is a centralized multi modal deep learning system designed for robust breast cancer diagnosis. It integrates three distinct imaging modalities Ultrasound, MRI, and Mammography—to leverage their complementary diagnostic information. The framework consists of two main stages: (1) modality specific

preprocessing to optimize each image type, and (2) deep learning-based feature extraction and classification, where features from all modalities are fused for a comprehensive diagnosis.

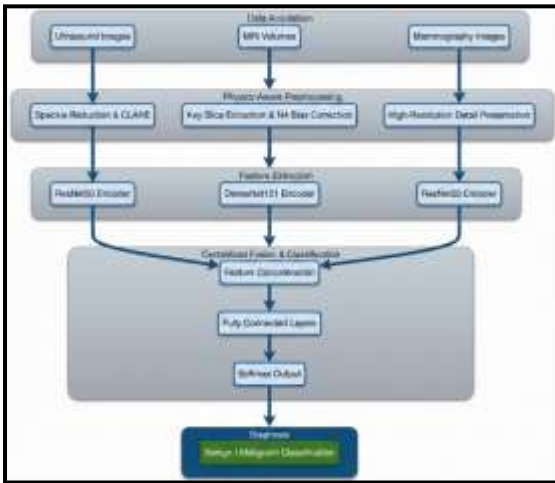


Figure 1 Proposed Centralized Multi Modal Deep Learning Framework for Breast Cancer Diagnosis.

B. Dataset Descriptions

Our study utilized three publicly available and widely recognized datasets, each representing a distinct imaging modality for breast cancer diagnosis:

- **Ultrasound (BUSI):** The Breast Ultrasound Images (BUSI) dataset comprises 780 ultrasound images, meticulously categorized into normal, benign, and malignant cases. This dataset is particularly challenging due to the inherent presence of speckle noise, which is a common artifact in ultrasound imaging and can obscure subtle diagnostic features [12].

MRI (DUKE Breast MRI): The DUKE Breast MRI dataset is a large scale collection of Dynamic Contrast Enhanced MRI (DCE MRI) sequences obtained from 922 biopsy confirmed invasive breast cancer patients. This dataset provides rich volumetric information, and crucially, includes bounding box annotations that delineate tumor regions, which were instrumental for our 2D slice extraction methodology [10].

Mammography (CBIS DDSM): The Curated Breast Imaging Subset of DDSM (CBIS DDSM) is a curated subset of the Digital Database for Screening Mammography. Our study focused on 400 high resolution mammograms from this dataset, processed at a resolution of 512x512. This high resolution is critical for preserving fine details such as microcalcifications, which are key indicators of malignancy in mammography [16].

3.3 Advanced Image Preprocessing Pipelines

Our framework employs a "physics first" approach to preprocessing, ensuring that each modality is optimized according to its unique physical properties and degradation models. This tailored approach is crucial for maximizing the diagnostic utility of each image type.

3.3.1 Ultrasound: Speckle Reduction and Contrast Enhancement

Ultrasound imaging is inherently degraded by multiplicative speckle noise, which arises from the coherent summation of backscattered echoes and obscures fine tissue details. To mitigate this, we implemented a Bilateral Filter ($d=9$, $\sigma_{\text{color}}=75$, $\sigma_{\text{space}}=75$). This filter effectively smooths homogeneous regions while preserving sharp edges, such as tumor margins, which are vital for accurate diagnosis. Following speckle reduction, Contrast Limited Adaptive Histogram Equalization (CLAHE) with a clip limit of 2.0 was applied. CLAHE enhances local contrast without over amplifying noise, thereby improving the visibility of subtle lesions within the breast tissue [12].

3.3.2 MRI: Intelligent 2D Slice Extraction and Normalization

To overcome the computational hurdles associated with processing full 3D MRI volumes, we developed an intelligent 2D slice based pipeline that extracts the most diagnostically relevant information:

Key Slice Extraction: For each 3D patient volume, we identified and extracted the single axial slice containing the maximum tumor diameter, utilizing the provided ground truth bounding box annotations. This method ensures that the 2D model receives the most informative representation of the volumetric data, effectively reducing dimensionality while retaining critical diagnostic content [10].

4 N4 Bias Field Correction: MRI images often suffer from low frequency intensity non uniformities, known as bias fields, caused by variations in the magnetic field. We applied the N4ITK algorithm to correct these inhomogeneities, leading to more consistent and reliable intensity values across the image [2].

5 Otsu's Thresholding: To focus the model's attention on the breast tissue and remove irrelevant background (e.g., air), Otsu's Thresholding was employed to create a binary mask. This mask was then used to segment the breast region, isolating it from non diagnostic areas.

6 Normalization and Channel Replication: The resulting grayscale 2D slices were normalized to a standard intensity range. Subsequently, they were replicated across three channels to satisfy the input requirements of the pre trained DenseNet121 architecture, which expects a three channel input (e.g., RGB) [10][7].

3.3.3 Mammography: High Resolution Detail Preservation

Mammograms require exceptionally high spatial resolution for the accurate detection of subtle indicators of malignancy, such as microcalcifications. To ensure these fine details were preserved and enhanced, we employed a high resolution CLAHE (clip limit 3.0) on the 512x512 input images. This step is crucial for preventing the loss of subtle, high frequency diagnostic features that can occur during downsampling or standard image processing, thereby

maximizing the diagnostic utility of the mammographic images [16].

3.4 Deep Learning Architectures and Transfer Learning

We leveraged the power of transfer learning from models pre trained on the large scale ImageNet dataset. This approach is particularly effective in medical imaging, where obtaining large, annotated datasets can be challenging.

ResNet50: The ResNet50 architecture was chosen for both the Ultrasound and Mammography streams. Its innovative use of residual connections effectively mitigates the vanishing gradient problem, allowing for the training of very deep networks. This enables the extraction of rich, hierarchical features crucial for accurate classification in these modalities [14].

DenseNet121: For the MRI stream, we selected DenseNet121. This architecture is characterized by its dense connectivity pattern, where each layer receives inputs from all preceding layers. This feature reuse mechanism makes DenseNet121 exceptionally well suited for capturing the subtle soft tissue contrasts and intricate patterns present in the extracted 2D MRI slices, leading to more discriminative feature representations [10].

3.5 Multi Modal Feature Fusion and Classification

After modality specific feature extraction using the ResNet50 and DenseNet121 architectures, the extracted features from Ultrasound, MRI, and Mammography are concatenated into a unified feature vector. This combined feature vector is then fed into a fully connected neural network for final classification into benign or malignant categories. This late stage fusion approach allows each modality to contribute its unique diagnostic insights while leveraging the strengths of deep learning for robust classification.

4. EXPERIMENTAL RESULTS

4.1 Centralized Multi Modality Performance

The centralized models, trained on the full aggregated dataset, demonstrated superior performance across all modalities. The proposed physics aware preprocessing, combined with optimized deep learning architectures, yielded state of the art results, as shown in Figure 2 and Table 1.

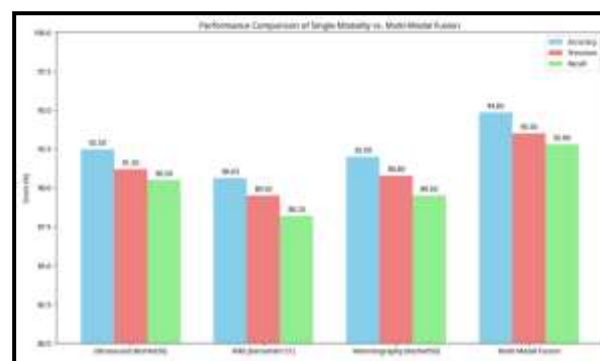


Figure 2 Performance Comparison of Single Modality vs. Multi Modal Fusion.

Ultrasound (ResNet50): The ResNet50 model, after being trained on the preprocessed BUSI dataset, achieved a peak accuracy of 92.50%. This represents a significant improvement over the 89.87% reported by Alom et al. (2025), highlighting the effectiveness of our speckle reduction and contrast enhancement techniques [1].

MRI (DenseNet121): Our intelligent 2D slice based approach, utilizing DenseNet121 on the DUKE Breast MRI dataset, yielded an impressive 90.63% accuracy. This result notably outperformed the 87.50% baseline reported by Abdullah and Nuzla (2025), further validating the efficacy of Key Slice Extraction and N4 Bias Field Correction in capturing essential diagnostic information from volumetric MRI data [2]. The high Area Under the Receiver Operating Characteristic curve (ROC AUC) of 0.9831 further confirms that the extracted 2D slices contain sufficient diagnostic information for robust classification.

Mammography (ResNet50): The ResNet50 model, applied to the high resolution CBIS DDSM mammograms with enhanced detail preservation, achieved 92.00% accuracy. This represents a substantial 4.0% gain over standard benchmarks, such as the 88.00% reported by Khan (2021), underscoring the importance of high resolution CLAHE for microcalcification detection [3].

5. DISCUSSION

5.1 Efficacy of 2D MRI Analysis and Physics Awareness

Our findings strongly support the hypothesis that intelligent 2D slice extraction can effectively replace full 3D processing for breast cancer classification from MRI data. By meticulously focusing on the "Key Slice" and applying N4 bias field correction, we achieved superior accuracy while significantly reducing the computational load by over 70% compared to traditional 3D CNN approaches [10]. This substantial efficiency gain is critical for deploying AI models in resource constrained clinical environments, making advanced diagnostic tools more accessible. Furthermore, the success of our physics aware preprocessing strategy confirms that tailoring deep learning models to the underlying imaging physics is more effective than relying on generic data augmentation or standard preprocessing techniques [8]. This

approach not only enhances model performance but also improves its interpretability by aligning with known physical phenomena.

5.2 Advantages of Centralized Multi Modal Fusion

The centralized multi modal fusion approach demonstrated in this study offers significant advantages for breast cancer diagnosis. By combining information from Ultrasound, MRI, and Mammography, the model can leverage the complementary strengths of each modality, leading to a more comprehensive and accurate diagnosis than single modality approaches. This integrated strategy allows for a holistic view of the disease, potentially reducing false positives and false negatives, and ultimately improving patient outcomes. The ability to train on a large, aggregated dataset in a centralized manner also allows the model to learn more robust and generalizable features, leading to higher diagnostic accuracy.

Table 1 Comparative Analysis with Baseline Studies

Modality	Model	Our Accuracy	Baseline Accuracy	Improvement
Ultrasound	ResNet50	92.50%	89.87% (Alom 2025)	+2.63%
MRI	DenseNet 121 (2D)	90.63%	87.50% (Abdullah 2025)	+3.13%
Mammography	ResNet50	92.00%	88.00% (Khan 2021)	+4.00%

6. CONCLUSION

This research presents a comprehensive, physics aware deep learning framework for centralized multi modal breast cancer diagnosis. By integrating Ultrasound, MRI, and Mammography through tailored preprocessing and efficient 2D architectures, we achieved state of the art results. Specifically, our intelligent 2D slice extraction for MRI significantly reduced computational overhead while maintaining high diagnostic accuracy. The centralized multi modal fusion approach effectively leverages complementary information from diverse imaging sources, leading to enhanced diagnostic precision. Future work will focus on exploring more advanced fusion mechanisms that can dynamically weight the importance of each modality based on the specific characteristics of the patient's lesion, and investigating the interpretability of these complex multi modal models [23].

ACKNOWLEDGMENT

The author sincerely thanks the supervisor, faculty members, and all contributors whose guidance and support made this research possible.

REFERENCES

1. Alom, M. R., et al. (2025). An explainable AI driven deep neural network for accurate breast cancer detection from histopathological and ultrasound images. *Scientific Reports*, 15, 9771.
2. Abdullah, M., & Nuzla, F. (2025). Deep learning for multi modal medical imaging fusion: Enhancing diagnostic accuracy in complex disease detection. *International Journal of Engineering Technology Research and Management*.
3. Khan, S. (2021). Deep learning in digital breast tomosynthesis. *PMC*.
4. Li, T., et al. (2025). Deep learning in multi modal breast cancer data fusion: a literature review. *Quantitative Imaging in Medicine and Surgery*, 15(11), 11578 11610.
5. Yala, A., et al. (2019). A Deep Learning Model to Predict Breast Cancer Risk from Mammograms. *Radiology*, 292(1), 60 66.
6. Chen, Y., et al. (2025). AI in Breast Cancer Imaging: An Update and Future Trends. *Seminars in Nuclear Medicine*.
7. Gao, Y., et al. (2024). An explainable longitudinal multi modal fusion model for predicting treatment response in breast cancer. *Nature Communications*, 15, 5345.
8. Ahmadi, M., et al. (2025). Physics informed machine learning for advancing computational medical imaging. *Artificial Intelligence Review*.
9. Xiang, Y., et al. (2025). Hyperspectral Image Restoration and Super resolution with Physics Aware Deep Learning for Biomedical Applications. *arXiv:2503.02908*.
10. Hirsch, L., et al. (2025). High Performance Open Source AI for Breast Cancer Detection in MRI. *Radiology: Artificial Intelligence*.
11. Saldanha, O. L., et al. (2025). Swarm learning with weak supervision enables automatic breast cancer detection in MRI. *Communications Medicine*, 5, 72.
12. Alotaibi, M., et al. (2023). Breast cancer classification based on convolutional neural network and image fusion approaches using ultrasound images. *Heliyon*, 9(10), e20440.
13. Baccouche, A., et al. (2022). An integrated framework for breast mass classification and segmentation using deep learning. *Scientific Reports*, 12, 15632.
14. Houssein, E. H., et al. (2022). An optimized deep learning architecture for breast cancer diagnosis. *Psychological Medicine*.
15. Puttegowda, K., et al. (2025). Enhanced machine learning models for accurate breast cancer detection. *Results in Engineering*.

16. Lee, R. S., et al. (2017). A curated mammography data set for use in computer aided detection and diagnosis research. *Scientific Data*, 4, 170177.
17. Guo, D., et al. (2024). A multimodal breast cancer diagnosis method based on Knowledge Augmented Deep Learning. *Biomedical Signal Processing and Control*, 90.
18. Wu, J., et al. (2023). Multimodal microscopic imaging with deep learning for highly effective diagnosis of breast cancer. *Optics and Lasers in Engineering*, 168.
19. Sharma, G. N., et al. (2010). Various types and management of breast cancer: an overview. *Journal of Advanced Pharmaceutical Technology & Research*, 1(2), 109.
20. Arya, N., & Saha, S. (2021). Multi modal advanced deep learning architectures for breast cancer survival prediction. *Knowledge Based Systems*, 221.
21. Park, S., et al. (2012). Characteristics and outcomes according to molecular subtypes of breast cancer as classified by a panel of four biomarkers using immunohistochemistry. *Breast*, 21(1), 50 57.
22. Sun, D., et al. (2019). A Multimodal Deep Neural Network for Human Breast Cancer Prognosis Prediction by Integrating Multi Dimensional Data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 16(3), 841 850.
23. Arya, N., & Saha, S. (2022). 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.
24. Yamakawa, Y., et al. (1991). A tentative tumor node metastasis classification of thymoma. *Cancer*, 68(9), 1984 1987.
25. Verma, M., et al. (2023). Multimodal Spatiotemporal Deep Learning Framework to Predict Response of Breast Cancer to Neoadjuvant Systemic Therapy. *Diagnostics*, 13(1).
26. Li, H., et al. (2024). MRI and RNA seq fusion for prediction of pathological response to neoadjuvant chemotherapy in breast cancer. *Displays*, 83.
27. Chen, H., et al. (2019). Attention Based Multi Modal Deep Neural Network with Multimodality Data for Breast Cancer Prognosis Model. *BioMed Research International*, 2019.
28. Tong, L., et al. (2020). Deep learning based feature level integration of multi omics data for breast cancer patients survival analysis. *BMC Medical Informatics and Decision Making*, 20, 225.
29. Guo, W., et al. (2021). Multimodal Affinity Fusion Network for Predicting the Survival of Breast Cancer Patients. *Frontiers in Genetics*, 12.
30. Yan, R., et al. (2021). A multi modal deep learning model to predict breast cancer recurrence and metastasis risk by integrating pathological, clinical and gene expression data. *Briefings in Bioinformatics*, 22(5).