Deep Learning Classification of Inflammatory Breast Cancer Using Multiparametric MRI: A Multi-Sequence Analysis

S. Ramezani[1], H. Le-Petross[2], M. Naser[1], R. V. Young[1], K. T. Morris[1], Kai M.[3], M.C. Stauder[4], B. Lim[3], A. Lucci[5], MDACC-IBC[6], C.D. Fuller[1], W. A. Woodward[4], C. R. Goodman[4]

[1]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, [2]Department of Breast Imaging, The University of Texas MD Anderson Cancer Center, Houston, TX, [3]Department of Breast Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, [4]Department of Breast Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, [5]Department of Breast Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX,
[6]MDACC Inflammatory Breast Cancer (MDACC-IBC) Team Authorship Contributors: Rachel Layman[1], Bora Lim[1], Azadeh Nasrazadani[1], Sadia Saleem[1], Vicente Valero[1], Michael C. Stauder[2], Wendy A. Woodward[2], Anthony Lucci[3], Susie X. Sun[3], Gary J. Whitman[4], Miral Patel[4], Huong Le-Petross[4], Yang Lu[5], Angela Marx[1], Angela Alexander[1], Chasity Yajima[1], Megumi Kai[1], Lily Villarreal[1], Heather Lopez[1].
[1]Department of Breast Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, [2]Department of Breast Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, [3]Department of Breast Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, [4]Department of Breast Imaging, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, [5]Department of Nuclear Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

THE UNIVERSITY OF TEXAS MD Anderson Cancer Center — Making Cancer History®

Background and Significance

- IBC is an aggressive and fatal disease: 2-4% of all breast cancer cases - 10% of all breast cancer mortalities in the US, affecting over 18,000 women.
- IBC has worse survival compared to locally advanced breast cancer (LABC) - 2.9 years vs 6.4 years, with cancer cells blocking lymphatic vessels in the skin.
- IBC is always classified as T4d and assigned at least stage IIIB at diagnosis.
- Diagnosis relies on subjective clinical criteria (rapid onset of erythema or edema involving ≥1/3 of the breast), leading to under-diagnosis and delayed treatment.
- An objective decision support system could reduce under-diagnosis and improve treatment decisions.
- Deep learning (DL) methods have shown promise in oncologic imaging, but IBC's unique radiologic features (diffuse patterns, edema) remain minimally explored.
- Multiparametric MRI (mpMRI) is the most sensitive imaging modality for detecting IBC features (global skin thickening, chest wall edema, non-mass enhancement).
- Large-scale annotated mpMRI datasets for IBC are limited, hindering development of robust AI diagnostic tools.
- Our approach leverages the world's largest clinically annotated IBC mpMRI registry at MD Anderson Cancer Center.

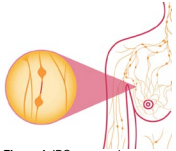


Figure 1. IBC occurs when cancer cells block lymphatic vessels in the skin of the breast, leading to rapid swelling, redness, and thickening.

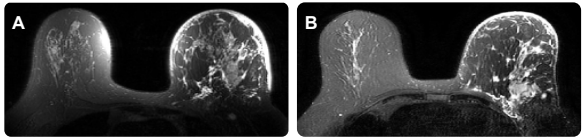


Figure 2. Axial T2-weighted post-contrast MRI images from two different patients showing (**A**) cT4b and (**B**) cT4d disease. Both demonstrate prominent skin thickening, highlighting the overlapping imaging features that contribute to diagnostic difficulty in distinguishing IBC from locally advanced breast cancer.

Problem Statement: Distinguishing IBC (cT4d) from locally advanced non-inflammatory breast cancer with extensive skin involvement remains challenging.

Study Objective: To develop and validate a DL-based automated classification system for IBC using mpMRI that provides objective, quantitative diagnostic support to clinicians.

Patient Cohort and Data

- Institutional patient registries at MD Anderson Cancer Center were used for patient identification and mpMRI data curation. Two independent mpMRI cohorts were evaluated: an IBC vs. Non-IBC cohort (T1NC N=741, DWI N=607, DCE N=478, T2 N=264) and a separate IBC vs. LABC cohort (T1NC N=711, T2 N=600, DCE N=658, DWI N=451).
- The IBC vs. Non-IBC cohort was split 80-20 for training and testing based on MRI sequence availability, resulting in variable sample sizes across T1NC, T2, DCE, and DWI.
- For the IBC vs. LABC cohort, the testing set included 40 patients with all four MRI sequences, and the training set was class-balanced by subsampling IBC cases to match the number of LABC patients.

Model Development and Analysis

Image Preprocessing

- MRI images were cropped at the midline to create single-breast (ipsilateral) images and bilateral full-breast images for comparative analysis. Images were resized to 144³ voxels for standardized input dimensions. Intensity normalization ensured consistency across different acquisition protocols. Data augmentation included random rotations, flips, and affine.

Model Architecture & Training

- DenseNet121-based DL architecture was employed for automated binary classification of IBC.
- Separate models were developed for each MRI sequence (DWI, DCE, T2).
- Five-fold cross-validation was performed to assess model robustness and minimize overfitting.

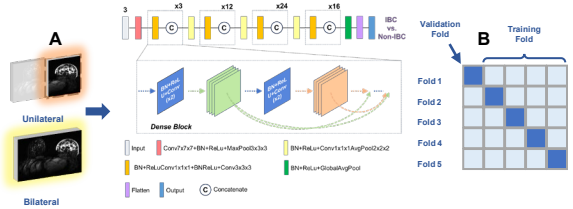


Figure 3. Deep learning classification framework. (**A**) DenseNet121 architecture with dense blocks and skip connections for automated IBC detection from unilateral or bilateral breast MRI. (**B**) Five-fold cross-validation with stage-aware stratified splitting ensuring robust model evaluation across all patient subgroups.

Model Evaluation

- Performance was evaluated using area under the receiver operating characteristic curve (AUC-ROC). Ensemble predictions were generated by combining outputs from all five folds.

Overall Classification Results

- Deep learning classification demonstrated strong discriminative ability across all MRI sequences for IBC vs. Non-IBC, with DWI and T2 achieving the highest ensemble performance (AUC=0.96), followed by T1 (AUC=0.89) and DCE (AUC=0.80).
- Models demonstrated robust cross-validation stability with minimal variance across folds, and ensemble predictions consistently outperformed individual folds.

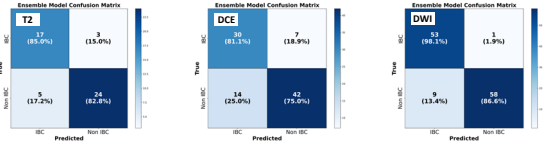


Figure 4. Ensemble model confusion matrices for unilateral breast MRI classification. DWI achieved the best overall performance with 98.1% IBC sensitivity and 86.6% non-IBC specificity, followed by T2c (85.0%/82.8%) and DCE (81.1%/75.0%).

Overall Model Performance

IBC vs. Non-IBC: DWI achieved the highest performance with ensemble AUC=0.96, followed by T2 (unilateral AUC=0.94, bilateral AUC=0.96) and DCE (unilateral AUC=0.85, bilateral AUC=0.80). High accuracy was achieved for early-stage disease (cT1-cT3: 92-97%) and classic IBC (cT4d: 94-95%). Distinguishing locally advanced breast cancer (cT4b) from IBC (cT4d) remained challenging (40-42% accuracy).

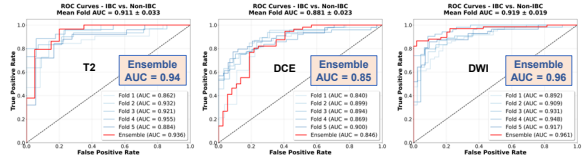


Figure 5. ROC curves for IBC vs. non-IBC classification across MRI sequences using unilateral (top) and bilateral (bottom) imaging. Individual fold performance (blue) and ensemble predictions (red) are shown. DWI achieved the highest discriminative performance (ensemble AUC=0.96) in both imaging approaches.

IBC vs. LABC: DCE showed the highest performance with an ensemble AUC=0.86, followed by T2 (AUC=0.82), T1 non-contrast (AUC=0.74), and DWI (AUC=0.76).

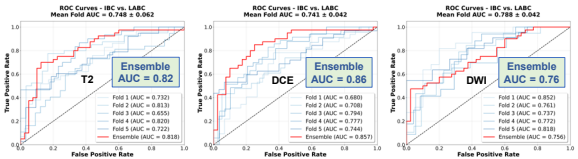


Figure 6. ROC curves for IBC vs. LABC classification across mpMRI sequences. DCE achieved the highest ensemble AUC (0.86), followed by T2 (0.82) and DWI (0.76).

Conclusions and Future Directions

- mpMRI deep learning enables strong IBC vs. non-IBC discrimination and may improve diagnostic accuracy and reduce delays.
- Lower performance for IBC vs. LABC aligns with known clinical diagnostic challenges.
- Limited cohort size underscores the need for multi-institutional validation.
- **Future work:** multi-sequence integration with demographics, clinicopathologic data, medical photography, and Komen Clinical Score.

Acknowledgement

This work was *supported by a fellowship from the Gulf Coast Consortia, on the NLM Training Program in Biomedical Informatics and Data Science T15 LM007093*