

MACHINE LEARNING-DRIVEN ARTIFICIAL MRI SYNTHESIS FROM UNPAIRED MULTIMODAL DATA FOR ENHANCED BREAST CANCER DIAGNOSIS

Aarthi Muthukumar, Amanda Brooks, Isain Zapata
Rocky Vista University, College of Osteopathic Medicine

Corresponding Author: Aarthi Muthukumar
Tel: +1 (435) 222-1236
Email: aarthi.muthukumar@ut.rvu.edu
doi:10.34107/UKKK6693.079

ABSTRACT

Breast cancer remains a leading cause of cancer-related morbidity in women worldwide, underscoring the urgent need for innovative diagnostic solutions. Early detection is critical for improving outcomes, yet existing imaging modalities — Digital Breast Tomography (DBT), magnetic resonance imaging (MRI), and ultrasounds (US)—each possess unique strengths and limitations. Advances in machine learning, particularly deep learning with Convolutional Neural Networks (CNNs), have demonstrated exceptional potential in medical imaging analysis. However, these methods are predominantly constrained to single-modality datasets due to the challenge of obtaining paired multimodal imaging data—a prerequisite for many supervised learning approaches. To address these limitations, this project proposes a novel computational framework for generating synthetic MRI data from DBT and US images. By utilizing unpaired image translation techniques, the model learns to synthesize high-fidelity MRI representations directly from DBT and US inputs. We capitalize on the fact that DBTs, ultrasound, and MRI images effectively capture the same anatomical structure. By leveraging this assumption, we can extract features from a shared latent space, eliminating the need for pixel-paired images across different modalities. The synthetic MRI data generated by this approach can serve as a bridge between imaging modalities, enabling multimodal training pipelines that harness the strengths of MG, DBT, and MRI in tandem. This allows us to address the scarcity of paired multimodal data, laying the foundation for new diagnostic paradigms that combine technological innovation with clinical accessibility.

Keywords: Breast cancer, convolutional neural networks, medical imaging, artificial image synthesis

INTRODUCTION

The rapid advancement of artificial intelligence in healthcare has transformed medical diagnosis and treatment planning, offering unprecedented opportunities to improve patient outcomes. In oncology, particularly breast cancer detection and diagnosis, machine learning algorithms have demonstrated remarkable potential in analyzing medical images with accuracy that sometimes rivals or exceeds human expertise. These technological innovations promise to enhance early detection rates, reduce false positives, and save lives. However, despite these promising developments, significant challenges remain in translating theoretical capabilities into practical clinical applications.

The primary limiting factor in developing effective machine learning algorithms is data availability. This constraint is particularly pronounced in medical imaging, where data access is restricted by patient privacy regulations and rare disease occurrences. The quality and quantity of training data directly determine the performance of machine learning systems, leading to a problematic trend: research efforts often concentrate on medical conditions with abundant imaging data, such as common cancers, while rare conditions remain underserved. Data augmentation is a critical strategy in enhancing the performance of deep learning models, particularly in image classification tasks. Mikołajczyk and Grochowski emphasize the significance of data augmentation techniques in improving the