Enhancing BI-RADS Classification in Digital Mammography: A Vision-Language Foundation Model Approach using Mammo-CLIP on the VinDr-Mammo Dataset

Chutimon Nunthasena[1], Siriporn Tubtim[1], Atchara Namburi[1,*]
[1]Department of Computer and Information Science, Science and Engineering Faculty, Kasetsart University, Sakon Nakhon, Thailand

**Corresponding Author Email: csnarn@ku.ac.th*

Abstract

Screening mammography is a critical tool for early breast cancer detection, yet the interpretation of mammograms remains challenging due to the high resolution of images and the complexity of lesion characteristics. Traditional Convolutional Neural Networks (CNNs) often face significant bottlenecks, including the loss of diagnostic details during image downsampling, high annotation costs for pixel-level labeling, and limited robustness across different clinical domains.

This study proposes a novel approach by utilizing Mammo-CLIP, a vision-language foundation model, to improve the accuracy and robustness of Breast Imaging-Reporting and Data System (BI-RADS) classification. By leveraging the VinDr-Mammo dataset, which contains 5,000 high-quality full-field digital mammography (FFDM) cases, this research addresses the data scarcity challenge through weak supervision from radiology reports and cross-modal alignment. Unlike conventional models, the proposed framework incorporates multi-view supervision to fuse features from both CC and MLO views, mimicking the diagnostic process of radiologists.

The expected results aim to demonstrate that the Mammo-CLIP architecture can achieve superior performance in BI-RADS categorization and lesion localization compared to traditional CNN baselines, particularly in terms of data efficiency and domain robustness. This work contributes to the development of more reliable computer-aided diagnosis (CADx) systems that can assist radiologists in reducing workload while maintaining high diagnostic standards.

Keyword: Breast Cancer, Mammography, BI-RADS, Vision-Language Model, Mammo-CLIP, Deep Learning, VinDr-Mammo.