# Explainable Deep Learning for Breast Cancer Classification and Localization


## Abstract

Breast cancer is a leading cause of cancer mortality among women, making early detection crucial. This paper explores the application of explainable artificial intelligence (XAI) in breast cancer diagnosis using ultrasound (US) imaging. It compares the performance of several convolutional neural networks (CNNs), including AlexNet, VGG16, GoogLeNet, a standard CNN, and a novel VMC-NET, in classifying breast tumours as benign or malignant. The study employs class activation mapping (CAM) to provide visual explanations for the model's predictions, enhancing clinical trust and interpretability.


## Introduction

Breast ultrasound (US) imaging is a non-invasive method for detecting structural abnormalities such as tumours, particularly useful in younger women with denser glandular tissue. Artificial intelligence (AI) has shown promise in improving breast cancer classification and diagnosis. This study aims to enhance the application of deep learning (DL) in clinical settings by integrating explainability, allowing clinicians to understand the reasoning behind a model's diagnosis. The paper focuses on employing convolutional neural networks (CNNs) and class activation mapping (CAM) to achieve both high accuracy and visual explainability.

## Methodology

The methodology involves several key steps: data acquisition and preprocessing, CNN model selection and training, and CAM application for explainability. The dataset, consisting of Ultra-sound images, was augmented to increase the number of samples and improve model generalization. The CNN models used include AlexNet, VGG16, GoogLeNet, a standard CNN, and a novel VMC-NET designed by the authors. The VMC-NET architecture includes convolutional layers, pooling layers, flatten layers, dropout layers, and dense layers. CAM is applied to generate heatmaps highlighting regions of interest (ROIs) that influenced the model's classification. The models are evaluated using metrics such as loss, accuracy, precision, recall, F-measure, and AUC.

## Experimental Analysis and Results

The dataset used in this study consists of PNG images with a resolution of 224x224 pixels. The dataset was divided into training, validation, and testing sets. Several experiments were conducted, varying parameters such as epochs, batch size, learning rate, and image size to optimize performance. The results showed variations in performance across the different CNN architectures, with some models achieving high accuracy and AUC scores under specific parameter configurations. Standard_CNN has achieved 92.81% accuracy. VMC_NET has outperformed other models in balancing accuracy and explainability with loss and accuracy values of 0.2038 and 0.9345 respectively.

## Conclusion

The study demonstrates the feasibility of using deep learning models for breast cancer classification with integrated explainability through CAM. The visual explanations provided by CAM can help clinicians understand the rationale behind the model's predictions, enhancing trust and acceptance in clinical practice. The comparative analysis of different CNN architectures provides insights into their suitability for this task, highlighting the importance of model selection and parameter tuning.

## References

*Explainable Deep Learning for Breast Cancer Classification and Localization*. ACM Digital Library. DOI:10.1145/3702237