

HYBRID CNN-TRANSFORMER FRAMEWORK FOR ACCURATE BREAST TUMOR DETECTION FROM ULTRASOUND IMAGE

Revathi.R¹, Aarthi.S², Sessa.P³, Swathi.R⁴, Thangam.M⁵

Department of Artificial Intelligence and Data Science

Bharathiyar Institute of Engineering for Women (Autonomous)

Abstract -Breast cancer is one of the most prevalent and life-threatening diseases among women worldwide. Early and accurate detection significantly improves patient survival rates. Ultrasound imaging is widely used for breast tumor diagnosis due to its safety, non-invasiveness, and cost-effectiveness. However, manual interpretation of ultrasound images depends heavily on radiologists' expertise and may result in diagnostic variability. In this paper, a Hybrid CNN-Transformer framework is proposed for accurate breast tumor detection from ultrasound images. The Convolutional Neural Network (CNN) extracts local spatial features, while the Transformer module captures global contextual relationships through attention mechanisms. The integration of these architectures enhances feature representation and improves classification performance. The proposed model classifies ultrasound images into benign, malignant, and normal categories. Experimental evaluation demonstrates improved accuracy and robustness compared to conventional CNN-based approaches. The system is further implemented as a web-based application to facilitate real-time tumor detection and assist medical professionals in decision-making.

Key Words: Breastcancer, Ultrasound imaging, Convolutional Neural Network, Transformer, Hybrid deep learning, Medical image classification.

I. INTRODUCTION

Breast cancer is one of the leading causes of cancer-related deaths among women worldwide. Early detection significantly improves treatment success and survival rates. Ultrasound imaging is widely used in breast tumor diagnosis due to its safety, affordability, and effectiveness in detecting abnormalities in dense breast tissues. However, ultrasound images often contain speckle noise and low contrast, making accurate identification challenging.

Deep learning techniques have shown promising results in medical image analysis. Convolutional Neural Networks (CNNs) effectively extract spatial features from images, but they mainly focus on local patterns.

Recently, Transformer-based models have demonstrated strong capability in capturing global contextual relationships using attention mechanisms. Combining these two approaches can enhance feature representation and improve classification performance. Therefore, this study proposes a hybrid deep learning framework to improve breast tumor detection from ultrasound images.

II. PROPOSED SYSTEM

The proposed system introduces a Hybrid CNN-Transformer framework for accurate breast tumor detection from ultrasound images. The main objective of this framework is to combine the advantages of Convolutional Neural Networks (CNN) and Transformer architectures to improve feature extraction and classification accuracy.

Initially, breast ultrasound images are provided as input to the system. These images undergo preprocessing operations such as resizing, normalization, and noise reduction to improve image quality and ensure consistent input dimensions for the model.

After preprocessing, the images are passed through a Convolutional Neural Network (CNN). The CNN layers are responsible for extracting local spatial features from the images. These features include edges, textures, shapes, and tumor boundaries, which help in identifying important visual patterns associated with breast tumors.

The feature maps obtained from the CNN are then forwarded to a Transformer encoder. The transformer utilizes a self-attention mechanism to capture global contextual relationships within the image. This mechanism enables the model to analyze the relationship between different regions of the image and identify

significant tumor-related patterns that may not be captured by CNN alone.

The features extracted by the CNN and the contextual information captured by the transformer are effectively combined to form a comprehensive feature representation. This hybrid architecture improves the model's ability to detect complex tumor structures and enhances classification performance.

Finally, a Softmax classifier is used to categorize the ultrasound images into three classes: Benign, Malignant, and Normal. The classifier produces the final prediction along with probability scores for each class.

To make the system practical and accessible, the trained model is integrated into a Flask-based web application. This web interface allows users to upload ultrasound images and receive real-time tumor prediction results along with confidence scores. The proposed system can assist radiologists in early detection and diagnosis of breast cancer, thereby improving clinical decision-making and patient outcomes

III. METHODOLOGIES

The proposed system follows a deep learning framework to accurately detect breast tumors from ultrasound images. The overall workflow consists of several stages including data collection, preprocessing, feature extraction, classification, and deployment.

1. Image Acquisition - Breast ultrasound images are collected from publicly available medical datasets. These images are used for both training and testing the deep learning model to detect tumor patterns.

2. Image Preprocessing - The collected images are preprocessed before feeding them into the model.

Images are resized to fixed dimensions to ensure uniformity in the dataset. Pixel values are normalized to improve model training performance. Noise reduction and contrast enhancement techniques are also applied to improve image quality.

3. Feature Extraction (CNN) - A Convolutional Neural Network (CNN) is used to extract important features from the ultrasound images. CNN layers automatically identify local spatial features such as edges, textures, and tumor patterns that are useful for classification.

4. Attention Mechanism (Transformer) - The feature maps extracted from CNN are passed to a Transformer encoder. The self-attention mechanism in the transformer

captures global contextual information and identifies important regions within the image.

5. Feature Fusion - Local features obtained from CNN and global features captured by the transformer are combined together. This fusion improves the representation of tumor characteristics and increases classification accuracy.

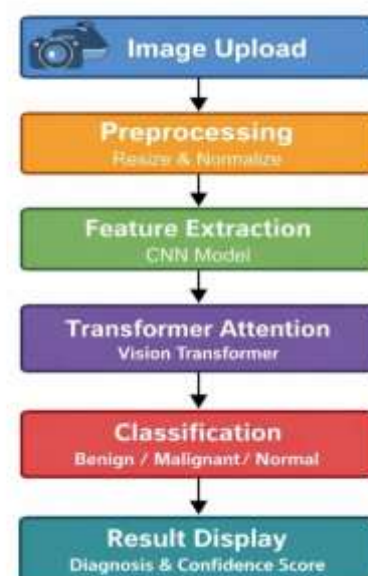
6. Classification - A Softmax classifier is applied to the extracted features to categorize the ultrasound images. The model classifies the images into three categories:

- Benign
- Malignant
- Normal

7. Model Training - The model is trained using the categorical Cross-Entropy loss function. The Adam optimizer is used to update model weights and improve learning performance during training.

8. Performance Evaluation - The performance of the proposed model is evaluated using standard evaluation metrics such as: Accuracy, Precision, Recall, F1-Score. These metrics help to measure the effectiveness of the tumor detection system.

9. Deployment - Finally, the trained model is deployed using a Flask web application. The system allows users to upload ultrasound images through a web interface and obtain real-time tumor prediction results along with confidence scores



Report generation the system generates a diagnostic report containing tumour prediction and confidence score, which can be saved or used for medical reference

Figure 3: Workflow of Hybrid CNN-Transformer model

IV. EXPERIMENT AND RESULT

Experimental Setup

The proposed Hybrid CNN–Transformer framework was implemented for breast tumor detection using ultrasound images. The experiments were conducted using a publicly available breast ultrasound dataset containing three classes: Benign, Malignant, and Normal.

Before training, the ultrasound images were preprocessed using techniques such as image resizing, normalization, and noise reduction to improve image quality and ensure consistent input dimensions for the model.

The dataset was divided into training and testing sets to evaluate the model performance effectively. The model was trained using standard deep learning optimization techniques to improve classification performance.

Evaluation Metrics

The performance of the proposed model was evaluated using commonly used classification metrics:

- **Accuracy** – Measures the overall correctness of the classification results.
- **Precision** – Indicates the proportion of correctly predicted positive cases.
- **Recall** – Measures the ability of the model to correctly detect tumor cases.
- **F1-Score** – Provides a balance between precision and recall.

Experimental Results

The experimental results show that the proposed Hybrid CNN–Transformer model achieved an overall accuracy of 97% in classifying breast ultrasound images.

The hybrid architecture combines CNN for local feature extraction and Transformer for capturing global contextual information, which improves the ability of the model to detect complex tumor patterns.

The trained model was integrated into a Flask-based web application, allowing users to upload ultrasound images and obtain real-time tumor prediction results. The system

also generates a diagnostic report containing tumor prediction and confidence score.

The experimental results demonstrate that the proposed system provides accurate and efficient breast tumor detection, which can assist medical professionals in early diagnosis and clinical decision-making.



Figure 4: Sign In Page of the System



Figure 7: Tumor Detection Result Page



Figure 5: Sign Up Page



Figure 8: Generated Diagnostic Report Page

Figure 6: Ultrasound Image Upload

Figure 9: Dashboard Interface of the System



Figure 10: User Account/Profile



V. CONCLUSION

The proposed hybrid CNN–Transformer framework for breast tumor detection from ultrasound images provides an efficient, reliable, and automated diagnosis system. By combining CNN’s local feature extraction with Transformer’s global attention mechanism, the model achieves improved classification accuracy and robust detection.

The web-based deployment allows real-time tumor prediction, assisting medical professionals in early breast cancer detection and reducing diagnostic errors.

Future enhancements include:

- Integration with hospital PACS systems
- Support for multi-modal imaging
- Mobile application development for wider accessibility.

VI. REFERENCES

1. Dosovitskiy et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” arXiv, 2020.
2. “BUSI Breast Ultrasound Dataset,” Public Dataset for Breast Ultrasound Image Analysis
3. World Health Organization, Breast Cancer Statistics Report, 2024.
4. Singh, S. P. Mishra, P. Singh, A. Srivastava, “VISNET: An Efficient Light Weighted Hybrid Model for Early Detection of Breast Tumor in Ultrasound Images using Vision Transformer and CNN,” Journal of Information Systems Engineering and Management, Vol. 9 No. 4s, 2024.
5. Q. He, Q. Yang, M. Xie, “HCTNet: A Hybrid CNN-Transformer Network for Breast Ultrasound Image Segmentation,” Computers in Biology and Medicine, 2023.
6. M. A. Rahman, “HyFormer-Net: A Synergistic CNN-Transformer with Multi-Scale Fusion for Breast Lesion Segmentation and Classification,” arXiv, Nov 2025.
7. G. Husain Abbas, S. U. Rehman, G. Bargshady, “A Hybrid CNN-Transformer Model for Multi

View Ultrasound Breast Tumor Classification,”
medRxiv, August 2025.

8. Y. Wei, “*Breast Ultrasound Image BI-RADS Classification Based on Vision Transformer,”* International Journal of Multiphysics, Vol. 18, No. 2, 2024.
9. B. Maistry, A. E. Ezugwu, “*Breast Cancer Detection and Diagnosis: A Comparative Study of Deep Learning Architectures,”* arXiv, May 2023.
10. J. Shi, A. Vakanski, M. Xian, et al., “*EMT-NET: Efficient Multitask Network for Breast Cancer Diagnosis,”* arXiv, 2022.
11. M. A. Rahman, “*HyFormer-Net: A Synergistic CNN-Transformer with Interpretable Multi-Scale Fusion for Breast Lesion Segmentation and Classification in Ultrasound Images,”* arXiv, Nov. 2025.
12. Mehmood, Y. Hu, S. H. Khan, “*A Novel Channel Boosted Residual CNN-Transformer with Regional-Boundary Learning for Breast Cancer Detection,”* arXiv, 2025.
13. Goodfellow, Y. Bengio, A. Press, 2016.