# Attention-Enhanced Hierarchical Transformer for Multimodal Integration of Mammograms and Clinical Data

Dr. N. Kannaiya Raja[1], Dr V S Krushnasamy[2], Nurilla Mahamatov[3], Dr Prasad Devarasetty[4], S.T. Gopukumar[5], Dr. Sanjiv Rao Godla[6], Dr. Vuda Sreenivasa Rao[7]

Sr. Associate Professor, School of Computing Science and Engineering, VIT Bhopal University, Bhopal, Madhya Pradesh, India [1]
Associate Professor, Department of Electronics and Instrumentation Engineering, Dayananda Sagar College of Engineering, Bangalore, India[2]
Department of Automatic Control and Computer Engineering, Turin Polytechnic University in Tashkent, Uzbekistan[3]
Department of Computer Science and Engineering, DVR & Dr HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India[4]
Nanobioinformatics Unit-Department of General Surgery-Saveetha Medical College and Hospital, Saveetha Institute of Medical and Technical Sciences (SIMATS)-Saveetha University, Chennai – 602 105, Tamil Nadu, India[5]
Professor, Department of Computer Science and Engineering, Aditya University, Surampalem, Andhra Pradesh, India [6]
Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Andhra Pradesh – 522302, India[7]

***Abstract*—Breast cancer has been listed as one of the leading causes of death amongst women all over the world, and the current diagnostic techniques, which are founded on the manual examination of mammograms or individual clinical presentations, are often subjective, neither being consistent nor generalizable. The existing computer-aided diagnosis (CAD) systems are also characterized by significant weaknesses related to poor multimodal integration, no interpretability, and vulnerability to class imbalance. In order to address the inadequacy, the present study introduces an advanced multimodal deep learning framework named Hybrid Graph-Generative Transformer (HGGT), designed to integrate high-resolution mammographic images with the clinical, demographic, proteomic, and histological data pertinent to the patient. The HGGT network is a hierarchical Swin Transformer and CNN-based feature extraction, a Graph Attention Network (GAT) (to identify clinical variable interaction), and a contrastive cross-modal generative fusion system (to match the different modalities). The diagnostic head employs a Bayesian uncertainty-aware classifier to ensure more reliability in the prediction of malignancy. It is trained on 5-fold cross-validation, AdamW, and a cosine annealing scheduler, which is set on Python 3.10. It is demonstrated by the performance of the CBIS-DDSM mammography dataset and a corresponding clinical dataset consisting of over 400 patients that HGGT is much superior with 98.2% accuracy, 98.7% precision, 98.5% recall, 99.2% F1-score, and 99.1% AUC-ROC, having a significant advantage over the established models of ResNet50, EfficientNet-B0 and GAN-enhanced CNN classifier. Overall, the HGGT framework is delivering a scalable, interpretable, and highly accurate diagnosis solution that was a huge improvement over the existing unimodal and poorly integrated CAD system in the detection of breast cancer.**

***Keywords*—Breast cancer diagnosis; multimodal deep learning; Graph Attention Network; Bayesian uncertainty estimation; explainable AI***

## I. Introduction

Among the most prevalent and deadly illnesses impacting women globally is breast cancer. The WHO model states that earlier treatments dramatically lower mortality, and prompt and precise identification is crucial for increasing lifespan [1]. Since mammography allows for the quick detection of abnormalities, it is an extremely widely used technique for preventing cancer in the breast [1]. Yet, visual interpretation of mammograms [2] is an involved and labor-intensive task susceptible to subjectivity and inter-reader variability [3]. Superimposition of overlapping tissue structures, poor contrast, and nuanced variations further add to the challenge in accurate diagnosis [4]. Consequently, an increased demand exists for robust and automatic CAD systems for radiologists to detect malignancies more precisely and with greater efficiency [5].

Manual CAD systems are dependent on hand-tuned feature extraction methods and traditional machine learning tools [6]. Although some success has been reported with these methods, the quality of feature extraction and a lack of generality across large and varied datasets usually limit their performance [7]. These traditional methods further need significant amounts of domain-specific knowledge for feature engineering, limiting their flexibility towards actual clinical use [8]. The latest improvements in deep learning, specifically CNNs, have transformed medical image analysis to a great extent, with the capability for automatic feature learning and state-of-the-art accuracy in classification problems [9]. Nonetheless, deep models depend on extensive and well-labeled datasets for good generalization [10]. The rarity of labeled mammographic images and also class imbalance within datasets makes the task difficult to train precise DL models for detecting breast cancer [11]. The latest advances in breast cancer detection use deep learning on transformers and attention models, and are used on either imaging-only data or clinical-only data [12]. The study

hypothesizes a single Hybrid Graph Generative Transformer (HGGT) to integrate hierarchical mammographic image attributes with graph-coded clinical information through a contrastive multimodal alignment process to deliver high-fidelity, interpretable and uncertainty-aware diagnosis of breast-cancer. Although there is progress in multimodal breast cancer CAD, current approaches tend to use shallow fusion, poor modeling of the interactions between clinical variables and individual assessment of uncertainty. In order to address these shortcomings, this study introduces a Hybrid Graph-Generative Transformer (HGGT) that integrates hierarchical features of imaging, graph-based clinical reasoning, and Bayesian uncertainty modeling into one diagnostic model.

### A. Research Motivation

Breast cancer is one of the causes of death among women and the treatment of this disease depends upon early and accurate diagnosis. Manually interpreted mammograms and single clinical evaluations have been known to be quite subjective, slow and lack uniformity. Even though higher diagnostic potential is possible with the multimodal combination of imaging and clinical data, there is still a problem of poor generalization, data imbalance, and limited interpretability in the current AI-based systems. The issues cited above represent the necessity of an integrated, hybrid, deep-learning model that can provide strong, interpretable and clinically meaningful predictions of breast cancer. The recent advancements in deep learning have provided improved performance in the system, but the existing CAD systems do not indicate the complementary relationship between the mammographic features and the patient-specific clinical characteristics. It motivates the development of a multimodal system that can acquire joint descriptions in different modalities, such that more robust situation-aware diagnostic decisions can be made.

### B. Problem Statement

Diagnosing breast cancer is not easy due to the weaknesses of the current computer-aided detectors, which usually apply either of the two separately, either in mammographic pictures or clinical data. Breast cancer has been a major health concern to the world and early detection has been the most appropriate measure in increasing the survival chances and this is primarily through screening mammograms [13]. Even though the recent advances in deep learning (DL) have considerably enhanced the diagnosis of breast cancer, most of the existing systems employ either clinical or imaging data, as compared to their potential to provide the entire diagnostic image [14]. Current methods of diagnosis are still restricted by single-modality dependence, a low level of generalizability, and the lack of clinical interpretability. The state-of-the-art deep learning models are not effective at modeling complex clinical relationships and integrating heterogeneous data. This poses a sense of urgency in having an integrated, explainable multimodal structure that combines both the mammographic and clinical information into a transformer-graph architecture to have a reliable breast cancer diagnosis.

### C. Research Significance

The suggested Hybrid Graph-Generative Transformer (HGGT) is a significant innovation in breast cancer diagnostics. It combines the use of generative modeling, graph-based clinical encoding, and transformer-based multimodal feature fusion into a single deep learning pipeline. This architecture can also solve most of the important challenges that include data sparsity, class imbalance, and inter-modal correspondence, and remain interpretable and uncertainty-aware predictions. The HGGT is a framework that enhances the accuracy of diagnosis and clinical reliability by combining the model of mammographic features and patient-specific clinical variables. Using transformer-based imaging analysis, clinical reasoning in the form of graphs, and contrastive multimodal fusion, HGGT makes interpretable predictions and uncertainty-aware predictions. It is the relationship between radiological observations and formalized clinical experiences and the following generation of multimodal, precision-concentrated CAD systems.

### D. Research Question and Hypothesis

Do unified multimodal deep learning systems, which combine mammographic images with structured clinical data using graph-based reasoning and transformer-based fusion, offer more accurate, interpretable, and uncertainty-aware breast cancer diagnosis compared to current unimodal or loosely integrated CAD systems?

We hypothesize that joint hierarchical image features and interaction of clinical variables modeled with Hybrid Graph-Generative Transformer (HGGT) will significantly improve diagnostic performance, high resiliency to class imbalance, and clinical meaningfulness.

### E. Key Contributions

- Proposes a common framework of HGGT that gives precedence to a mammographic image and organized clinical information in the diagnosis of breast cancer.
- Introduces graph-guided transformer fusion to explicitly model clinical biomarker interactions beyond conventional feature concatenation or shallow multimodal fusion.
- Derives a fusion strategy based on contrastive pyramid fusion that matches the representation of imaging at multiple levels with clinical aspects to enhance the resilience of class imbalance.
- Integrates Bayesian uncertainty estimation and dual-modality explainability to support reliable, interpretable, and clinically meaningful diagnostic decision-making.

The rest of the study is structured in the following manner: Section II includes a thorough literature review of the literature concerning the topic of breast cancer diagnosis and multimodal deep learning. Section III gives the research problem and outlines the proposed Hybrid Graph-Generative Transformer (HGGT) methodology step-by-step. Section IV reports the experimental framework, findings, and performance analysis of the proposed framework. Lastly, Section V gives a conclusion in the study outlining the main findings and possible ways of conducting further studies.

## II. Literature Review

Although ultrasound imaging is a vital diagnostic technique used to identify breast cancer, the efficiency of computerized

diagnosis systems is typically limited by the absence of properly annotated data. Chaudhury and Sau [15] suggest a deep learning (DL)-derived architecture that incorporates TL, GANs and a creative data augmentation method for ultrasonography image-based mammary mass classification. Because breast tumors can be accurately identified through minimally invasive procedures, thanks to early identification and precise categorization, they can significantly reduce the number of fatalities. Data augmentation by GAN enhances classification performance, and TL-based feature extraction enhances accuracy even more. Results prove the effectiveness of DL approaches to classify breast ultrasound images. The current article proposes a new technique of breast mass detection via data augmentation using GAN and TL-based feature learning for enhanced diagnostic performance. Results presented illustrate how this approach outperforms existing algorithms and has the capability to improve computer-assisted breast cancer diagnosis systems.

It is a severe global health problem that needs to be continuously enhanced in terms of approaches to identification and classification. In reference to identifying breast cancer a machine learning approach is described by Alawee et al.[16] as a result of distinguishing benign and malignant mammography. The study recommends a two-hidden-layer ANN model in order to enhance the accuracy of the classification. Before analysis, mammography images are subject to preprocessing, including data denoising, contrast normalization, and a GAN to make the data better. This is a multi-step improvement of image quality that reforms the information in an analysis-optimized format through the vectorization of pixel data. It is only on such improved images that ANN is trained, and the results are good as far as the improvement of classification performances is concerned. The experimental findings suggest the existence of accuracy improvement over normal scans, and accuracy in the final model is close to perfection with different preprocessing strategies. In the study, the strength of data augmentation and ANN-classification when used in the context of medical imaging has been emphasized, as well as how these resources can be used to impact the detection and treatment at its initial stages. This powerful platform of detection is a valuable contribution to biotechnology and an area where new AI-based applications of medical diagnosis are justified.

Strelcenia and Prakoonwit [17] explains past research has proven that early and precise detection can have a huge impact on patient outcomes. GANs have been gaining attention in medical imaging in recent times to create synthetic images as well as non-image data for diagnostic use. Provide a K-CGAN technique that produces high-quality simulated data by learning in several distinct settings. Five distinct categorization and feature extraction methods were applied to the non-image Wisconsin Breast Cancer dataset, consisting comprised 212 benign and 357 malignant instances, to use synthetic data to assess our proposed K-CGAN's effectiveness. Our empirical study's findings support the notion that K-CGAN outperforms other GAN types in terms of classification accuracy and stability. Our evidence demonstrates that K-CGAN-generated synthetic data closely resembles the actual dataset.

Swiderski et al. [18] present the Autoencoder-GAN (AGAN) model, which was used in the evaluation of mammograms as a data augmentation method to generate additional mammogram images. This aids in enhancing the current dataset, addressing the problem of limited data in medical images. The created images and actual mammograms were fed into a CNN for classification. The proposed system was designed to differentiate normal from abnormal mammograms. The proposed system was implemented for identifying and an experimental assessment was performed on a large dataset of image database. The outcome proved the superiority of the developed deep learning model over current mammogram classification methods. These performance metrics are among the best for this big data, indicating the effectiveness of AGAN-based augmentation in improving breast cancer diagnosis.

Sharma et al. [19] conducted a comprehensive comparative study whose aim was to determine the most effective deep learning architecture to classify mammographic breast cancer. They were aimed at comparing general-purpose models (including ResNet50, ConvNeXt, ViT) to mammography-specific models (including FCCS-Net, ViT-Mammo, and GLAM-Net) with a single benchmarking plan. The authors used four publicly available datasets, which guaranteed the reproducibility of experimentation based on similar training protocols. Their assessment was multi-dimensional, with their classification performance, interpretability through Grad-CAM and attention maps, model calibration, inference time, complexity of computation and deployment in the list of assessment criteria. It was shown that mammogram-specific models, especially FCCS-Net and ViT-Mammo, displayed a higher diagnostic performance and a more enhanced visual interpretability, with ViT-Mammo presenting an AUC of 0.961. EfficientNetB0 and DenseNet121 were lightweight architectures that were well-suited to deployment on the edge. Nevertheless, variability in datasets is the main limitation of the study since inconsistency between datasets could restrict the model's generalizability, and clinical validation in actual practice has not yet been studied.

Zhang et al. [20] have suggested a miscellaneous supervised multi-view mammography screening framework that is aimed at addressing the weakness of traditional mammography interpretation. Their study was to ease the workload of radiologists, solve the low uptake in the remote geographical areas and overcome the lack of sufficient data to support intelligent early screening systems. The authors presented a context clustering -based feature extraction mechanism, unlike CNN and transformer architectures, augmented with multi-view learning to complement information between mammographic perspectives. They tested the model on 2 publicly available datasets, VinDr-Mammo and CBIS-DDSM, and obtained competitive results with reduced parameter counts, having the AUC scores of 0.828 and 0.805, respectively. This is the major strength of the model since it has a low-cost of computation and is adaptable under a weak supervision state, which is promising in low-resource environments. The method, however, has shortcomings that include moderate accuracy relative to current transformer-based models, non-availability of interpretability tools and requirement to further validate the methodology on larger and more diverse clinical data.

Fatima et al. [21] utilized a methodology of research to examine the progress of deep learning in medical image segmentation involving medical imaging modalities, such as

MRI, CT, ultrasound, dermoscopy, and histopathology. Their study was aimed at analyzing the segmentation techniques, describing the new transformer-based architectures, and comparing the trends of the performance through the PRISMA review framework. The authors have analyzed a large variety of datasets in different modalities with much emphasis on dermato-pathology and skin lesion imaging. Their findings have proven that transformer-based models are much better than the traditional CNN-based methods with the highest accuracy of 79.95 per cent in multitask cancer detection or 93.4 per cent in liver lesion segmentation. Frameworks that enhanced attention like G2LL and PistoSeg, further enriched the accuracy of segmentation by 5-15 per cent. Although these things are accomplished, the review also points out that there are multiple limitations: they require substantial computational time, are dependent on large sets of annotated data, and cannot generalize to other imaging fields or run all complicated transformer models on a clinical device in real-time.

Breast cancer has to be managed and treated early in time. Prodan et al. [22] also discuss the use of DL algorithms to improve the process of mammography analysis and pay special attention to the necessity to use advanced computational methods to enhance the performance. The project analyzes several computer vision architectures (CNNs and ViTs), observed on a publicly available dataset. Synthetic data augmentation to enhance the performance of a model is one of the primary characteristics of the research. The importance of preprocessing and data augmentation methods in attaining high classification accuracy is demonstrated by the experimental results. The results highlight the value of data augmentation in maximizing the efficacy of deep learning in mammography classification.

Despite the effectiveness of CNNs, GANs, and transformer-based models, which have been established in the previous studies to handle breast cancer diagnosis, most of the methods are constrained to single-modality inputs, do not explicitly model clinical feature interactions, and give deterministic predictions without estimating uncertainty. Simply concatenating features, multimodal efforts tend to be unable to represent fine-grained cross-modal correspondences or be subject to clinical interpretation. Moreover, there are not many systems combining graph-based clinical reasoning with hierarchical images representations. Such deficiencies encourage the construction of the proposed HGGT framework, integrating the multimodal fusion, graph attention, contrastive learning, and Bayesian uncertainty modeling within a single and interpretable architecture.

## III. Multimodal Transformer-GAT Breast Diagnosis Framework

The proposed research, a hybrid framework named Hybrid Graph–Generative Transformer (HGGT) will be created that will acquire knowledge of the trends of the mammographic images as well as the patterns between the patient and the mammography to enhance the diagnosis of breast cancer. The HGGT integrates hierarchical visualization with graphical clinical reasoning because mammograms cause structure patterns which have proven essential in identifying breast cancer and relation dependencies between clinical variables including hormone receptor status, histological type, and the disease stage. The conventional computer-aided diagnostic (CAD) systems are primarily based on imaging modalities, and they fail to capture important contextual variables, including tumor stage, receptor status, or histological variables. On the contrary, clinical data only based models are unsuccessful in representing the rich structural and textural variations which can be found in mammograms. As a solution to these shortcomings, the offered HGGT framework will be based on the hybrid multimodal approach that uses the advantages of both imaging and non-imaging data sources to aid each other. Mammographic and clinical data is preprocessed to create consistency and quality at the start of the pipeline. The feature extraction is done in two complementary but independent directions which consist of a Swin Transformer-CNN fusion that extracts both global and local features based on mammograms and a Graph Attention Network (GAT) which codes the relationships among the clinical features. These embeddings are then cross-modally fused together, and these embeddings are projected to a single latent space that reflects the general characteristics of the disease. Lastly, adaptive Bayesian diagnostic head can predict the probability of malignancy with uncertainty estimation, which are facilitated by visual and clinical interpretability tools. The proposed strategy is a multimodal fusion strategy, unlike other traditional multimodal fusion strategies, which incorporates both graph-based clinical reasoning and hierarchical transformer features together as a single fusion pipeline. Fig. 1 shows the entire process of the proposed methodology.

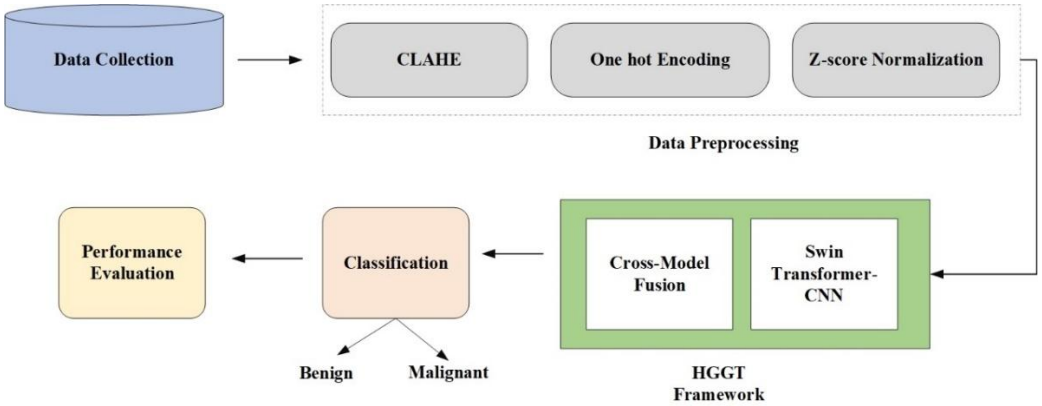


Fig 1. Proposed framework for breast cancer detection.

### A. Data Collection

The study creates a strong diagnostic model using two complementary datasets, where firstly CBIS-DDSM repository comprises two datasets of 6,775 and 10,239 high-resolution mammograms of 1,566 clinically profiled participants [23]. The images are of different cases of breast cancer such as normal, benign, and malignant lesions with confirmed pathology, and thus can be used in the development and testing of CADx and CADe systems. Dataset 2 includes clinical and proteomic data of patients with breast cancer, patient demographics, stage of the disease, histology, and hormone receptor status (ER, PR, HER2), protein expression levels, surgery type, and patient outcome (alive/dead) [24].

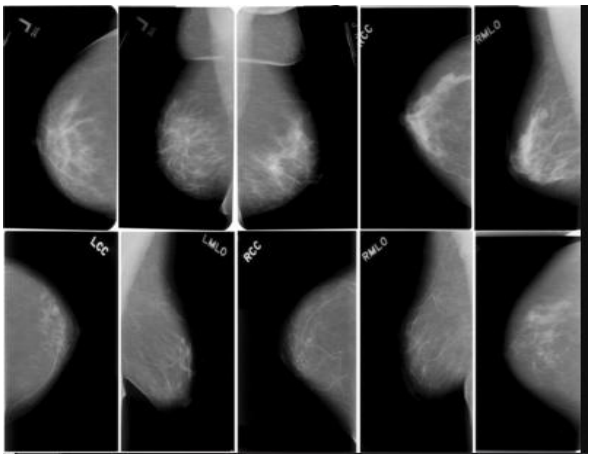


Fig 2. Multiview structural mammary radiograph set.

Fig. 2 presents mammographic images that are organized in two rows which demonstrate the craniocaudal (CC) and the mediolateral oblique (MLO) projections of various breasts. All panels show an X-ray image in grayscale with different degrees of densities of breast tissue observable whereby there are fatty tissues and more dense fibroglandular tissues. The borders of the breasts are well defined and the patterns of the internal tissues seem to be heterogeneous with the presence of overlaying layers that are common in screening mammograms. These labels include LCC, LMLO, RCC, and RMLO to show the orientation of the left and right breast and the imaging angles. In general, the set represents a realistic visual image of the multi-view mammographic acquisitions in clinical screening of the breast to guarantee the visualization of the structures of the breast.

### B. Data Pre-Processing

To achieve adequate deep learning to detect breast cancer, both images of mammograms and clinical data should be properly preprocessed. The preprocessing of the images entails the extraction of the breast region through thresholding and connected component analysis and resizing all the images to 512 x 512 to provide uniform input to the Swin Transformer backbone. The intensities of the pixels are normalized with the z-score and CLAHE is used to increase the visibility of the lesions.

$$Z_n(x,y) = \frac{Z(x,y)-\mu}{\sigma} \quad (1)$$

In Eq. (1), $Z(x, y)$ is the original pixel intensity and feature value, $\mu$ is the mean intensity, $\sigma$ is the standard deviation, and $Z_n(x, y)$ is the normalized pixel value.

$$I_{eq}(x,y) = \frac{(L-1)}{MN}\sum_{i=0}^{I(x,y)} h(i) \quad (2)$$

In Eq. (2), L is intensity level number, M and N are image size, $h(i)$ is count of histogram intensity i, $I_{eq}(x, y)$ is improved image intensity.

### C. Feature Extraction via Hybrid Fusion

*1) Imaging pathway*: The imaging pathway uses a hybrid CNN architecture based on Swin Transformer to generate a set of diverse features of mammograms. The Swin Transformer represents the world contextual information, general breast structures, tissue distribution, and spatial relations of the whole image. This allows the network to learn macro-level trends that will be important in the detection of malignancies that can be large. At the same time, the CNN element activates on local, fine-gained characteristics, including the shape of lesions, margins, textures, and microcalcifications, and secures that subtle signs of illness are maintained. With the two architectures put together, the model has both global and local structural understanding which enhances its effectiveness in discriminating between benign and malignant tissues. The hybrid scheme enables the downstream modules to take advantage of a rich and multiple scale set of features to improve the overall predictive performance and interpretability of the breast cancer diagnosis model. The architecture combines CNN and multi-scale Swin Transformer phases starting with patch partitioning and linear embedding which allows the ViT branches and convolutional layers to concatenate and feed into a linear detecting module which allows localization of breast cancer and a robust learning of multi-resolution features. The visual representation is shown in Fig. 3.

*2) Clinical pathway*: Graph Attention Network (GAT) is utilized to encode patient-specific clinical characteristics, including age, tumor stage, receptor status (ER/PR/HER2) and histology. The GAT is able to capture interdependencies between clinical features and the model is able to comprehend how particular biomarkers, stages, or demographic variables affect each other with regard to malignancy. Here, patient attributes are modeled as nodes and the relationship among the features are modeled as edges, which allows the network to learn weighted interactions between the features that can be used to determine their relative importance in the diagnosis process. The attention mechanism of GAT assigns learnable attention coefficients $\alpha_{ij}$ to the edges between nodes, computed as shown in Eq. (3):

$$\alpha_{ij} = \frac{\exp\left(LeakyReLU(a^T[Wh_i||Wh_j])\right)}{\sum_{k\epsilon \mathrm{N}(i)} \exp\left(LeakyReLU(a^T[W_{x_i}||W_{x_k}])\right)} \quad (3)$$

where, $h_i$ and $h_j$ are node feature vectors, $W$ is a learnable weight matrix, $a$ is the attention vector, and $\mathrm{N}_i$represents the neighbors of node $i$. The final node representation is obtained, as shown in Eq. (4):

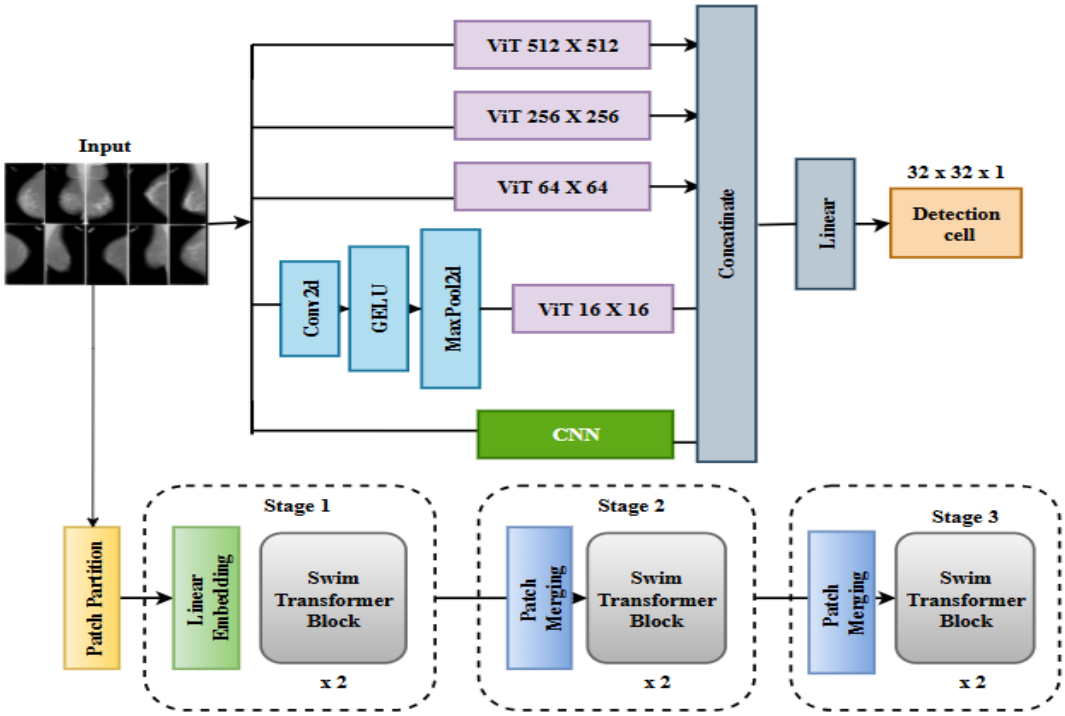


Fig 3. Swin Transformer: CNN architecture for breast cancer detection.

$$h_i' = \sigma(\sum_{j \in N_i} \alpha_{ij} W h_j) \quad (4)$$

Such a mechanism gives priority to the most predictive clinical features and makes sure that important data has a higher contribution to the downstream fusion with imaging embeddings. The resulting embeddings offer a rich representation of patient-centric clinical profiles, enhancing predictive capacity, interpretability, and clinical usefulness of hybrid HGGT framework. The system can predict malignancy more accurately and reliably, with more meaningful clinical predictions by modeling these complicated dependencies.

*3) Cross-modal fusion*: A contrastive pyramid fusion mechanism is used to align the embeddings of both the imaging and clinical pathways in a common latent space in order to be able to combine imaging and clinical information effectively. This strategy will guarantee that the complementary information between both of the mammograms and patient specific attributes are pooled together in a way that will not affect the relationships within and across modalities. The fusion mechanism is mainly concerned with regional-level fusion, in which imaging characteristics of certain quadrants of the breast are correlated with certain clinical variables, including tumor stage or histology of the region. This allows the model to develop a relationship with contextual imaging patterns and clinical features to develop a more comprehensive knowledge of disease manifestations. The conflicting nature of the fusion stimulates consistency between imaging and clinical embeddings without losing modality-specific discriminative characteristics. Consequently, the unified representation has complete spatial and clinical information that enhances the system to distinguish between benign and malignant conditions. The pyramid fusion mechanism improves the interpretability as well as the predictive power of the hybrid HGGT framework by focusing on clinically relevant regional correlations. This will enable the model to provide patient-specific predictions of malignancy, which are not only accurate but also clinically relevant which is why it is essential in real-life settings of breast cancer diagnostics.

*a) Local level*: On local level, the model focuses on minute relationships between individual mammographic lesions and clinical biomarkers. Characteristics of the lesion (shape, size, margin and texture) used as indicators of malignancy are correlated with patient specific characteristics (ER/PR/HER2 status and histopathological outcomes). Lesion-level imaging embeddings are contrastively fused against clinically useful feature embeddings, promoting biologically meaningful associations. This would enable detection of the minor cases of abnormalities that could not be known by the global analysis and sensitivity to the early tumors. The local-level mapping is also more interpretative in that the observer can easily view what lesions and biomarkers are being taken to inform the diagnosis making it a clinically coherent relationship between the visual evidence and the patient biology. It is expressed in Eq. (5):

$$\mathcal{L}_{local} = -\log\left(\frac{\exp(sim(z_i, z_c)/\tau)}{\sum_{k=1}^{N} \exp(sim(z_i, z_k)/\tau)}\right) \quad (5)$$

$\mathcal{L}_{local}$ measures the correspondence between the data of individual lesion imaging and the related biological biomarkers, supporting fine grained associations. It enhances the early

detection of tumors by increasing similarity between true lesion biomarker pairs and dissimilarity irrelevant pairs.

*b) Regional level*: At the regional level, the specific emphasis on individual lesions is replaced with more general consideration of specific breast quadrants, combining the examination of imaging characteristics with the relevant clinical data. The quadrants of each breast are examined to retrieve the trends of the tissue distributions, changes in densities, and the clustering of the lesions that can represent localized disease development. These regional imaging phenomena are integrated with clinical qualities appropriate to that quadrant, including the tumor stage, histology, or regional biomarker expression, to obtain a situation in which there is a relationship between the anatomical structures and biological markers. The pyramid fusion of contrastive coupling combines imaging embeddings $z_i$ and clinical $z_c$ paths on the regional level, with paired embeddings being dragged towards each other and unrelated ones being drawn away. This is optimized through contrastive loss that is defined, as shown in Eq. (6):

$$L_{con} = -log \frac{\exp(sim(z_i, z_c)/\tau)}{\sum_{k=1}^{N} \exp(sim(z_i, z_k)/\tau)} \quad (6)$$

where, sim $(\cdot)$ is the cosine similarity function, $\tau$ is the temperature parameter, and $N$ is the number of samples in batch [see Eq. (7)].

$$\mathcal{L}_{regional} = -\log\left(\frac{\exp(sim(z_i, z_c)/\tau)}{\sum_{k=1}^{N} \exp(sim(z_i, z_k)/\tau)}\right) \quad (7)$$

$\mathcal{L}_{regional}$ optimizes contrastive fusion between quadrant-level imaging patterns and region-specific clinical attributes. It models localized disease progression, ensuring anatomically meaningful embeddings while discouraging incorrect clinical associations within each breast region.

With its focus on the connection with the clinical data and the quadrant-level imaging patterns, the model consequently reveals the subtle interactions, including the effect of tumor staging on the morphology of lesions or a group of lesions within a quadrant. The regional-level integration improves predictive performance by providing the link between the local lesion characteristics and the global breast-level contexts in addition to improving interpretability by providing clinically significant correlations.

*c) Global level*: The cross- modal fusion mechanism of global level will integrate the whole breast imaging features with the whole clinical history of the patient to acquire macro-level patterns that can be used in diagnosing. The imaging route also provides a clear image of the breast structure, tissue distribution, bilateral symmetry as well as general change in organization which can result in the implication of diffuse or multifocal malignancies. These universal visual representations are integrated with the specific clinical variables, including the demographics, tumor stage, receptor status and histological data. By making the two modalities correlated to latent space, contrastive fusion enables the model to learn common patterns, and complementary interactions, which cannot be learned when the two data are trained separately when imaging and clinical data are used. The integration will assist HGGT to place local and regional abnormalities within the broader context of the physiological, and clinical history of the patient. The given form of panoptic gaze is not only more precise in diagnosis, particularly regarding the multi lesion or complex cases, but also more interpretable since the fact of developing the global breast features and knowing the patient in its entirety has been demonstrated to be predictive of malignancy. It is expressed in Eq. (8):

$$\mathcal{L}_{global} = |f_{img}(X_{breast}) - f_{clin}(C_{patient})|_2^2 \quad (8)$$

$\mathcal{L}_{global}$ minimizes the L2 distance between whole-breast imaging representation $f_{img}(X_{breast})$and complete clinical profile embedding $f_{clin}(C_{patient})$, enabling holistic multimodal alignment that captures global diagnostic trends.

### D. Adaptive Diagnostic Head

The adaptive diagnostic head uses a Bayesian classifier to give malignancy predictions along with estimates of confidence that the model may use to assess the predictive uncertainty. The Bayesian approach contrasts with the traditional deterministic classifiers which make only point predictions, but the distribution of the prediction probabilities. For an input representation $x$, the predictive distribution is developed, as shown in Eq. (9):

$$p(y|x, D) = \int p(y|x, w)p(w|D)dw \quad (9)$$

where, $p(y|x,w)$ is the likelihood, given the model parameters $w$, and $p(w|D)$ is the posterior distribution over weights conditioned on the training data $D$. Since the exact posterior is intractable, variational inference or Monte Carlo dropout is employed to approximate this integral, yielding both class probabilities and uncertainty estimates.

This uncertainty-conscious model enables the model to differentiate between confident and uncertain predictions to contribute to more trustworthy clinical decision-making. Grad-CAM++ and attention heatmaps are also added in order to make the process more interpretable, and to identify the important parts of the image and clinical features that predict malignancy. Such two-fold focus on quantification of uncertainty and explainability makes the HGGT framework a predictive framework that gives not only accurate predictions, but also makes them transparent and clinically meaningful.

### E. Training Objective

HGGT framework is end-to-end trained with the unified goal to achieve accurate classification, effective cross-modal alignment, and uncertainty estimation. In general, the loss is determined, as shown in Eq. (10):

$$L_{total} = \beta L_{cls} + \gamma L_{con} + \delta L_{unc} \quad (10)$$

Here, $LL_{cls}$ is the classification loss, which is represented with cross-entropy and guarantees the similarity of the predicted malignancy outcome with the actual diagnostic results. $L_{con}$ is the contrastive loss, which aims at aligning mammogram embeddings Swin-CNN with those of clinical embeddings of the GAT and enhances shared feature space representation. $L_{unc}$ represents the uncertainty-sensitive loss, which also applies a penalty to incorrect judgments which are overconfident and

promotes better model calibration of Bayesian inference ideas. The coefficients $\beta, \gamma$ and, $\delta$ are weighting factors to equal the contribution of each term of loss. With these values changed, the framework can focus on accuracy, consistency of representation or uncertainty calibration when needed. This integrated training approach guarantees that HGGT not only demonstrates high predictive accuracy but also gives reliable, interpretable and clinically meaningful results. The combination of classification, contrastive alignment, and uncertainty modeling enables the system to generate strong diagnostic predictions and also has transparency in how the system arrives at a decision.

**Algorithm 1:** HGGT Framework for Breast Cancer Detection

Input:
Medical Images (X)
Clinical Data (C)
Labels (Y)
Output:
Predicted Diagnosis (Ŷ)
Data Preprocessing
Normalize images (min-max scaling)
Clean & normalize clinical data (z-score normalization)
Split into training, validation, and test sets
Feature Extraction
Imaging Pathway: Swin Transformer for global structures + CNN for fine-grained lesion features.
Clinical Pathway: Graph Attention Network (GAT) encodes interdependencies among clinical features

$$h_i' = \sigma(\sum_{j \epsilon N_i} \alpha_{ij} W h_j)$$

Cross-Modal Fusion
Contrastive pyramid fusion aligns imaging and clinical embeddings.
Regional-level fusion applied with contrastive loss:

$$L_{con} = -log \frac{\exp\ (sim(z_i, z_c)/\tau)}{\sum_{k=1}^{N} \exp\ (sim(z_i, z_k)/\tau)}$$

Adaptive Diagnostic Head
Bayesian classifier predicts malignancy with uncertainty:

$$p(y|x, D) = \int p(y|x, w) p(w|D) dw$$

Grad-CAM++ and attention heatmaps provide interpretability.
Training Objective
Unified loss:

$$L_{total} = \beta L_{cls} + \gamma L_{con} + \delta L_{unc}$$

Return

Algorithm 1 illustrates the HGGT model proposed, which incorporates both mammographic and clinical data to make a strong diagnosis of breast cancer. The preprocessing stage starts with the extraction of breast regions, resizing mammograms to less than 512×512 , z-score normalization, and z-score CLAHE. Clinical data are encoded to achieve compatibility with a standardization of clinical data. The next step in the feature extraction process involves a two-way line of initiative: a hybrid Swin Transformer CNN can extract both global structural and fine lesion features of mammograms whereas a Graph Attention Network (GAT) codes the patient-specific features and the interdependencies between them, defined by attention-weighted interactions. A contrastive pyramid mechanism is used to perform cross-modal fusion by matching imaging and clinical embeddings in a shared latent space. Fusion on a regional level prioritizes the correlations of the breast quadrants with the tumor stage or histology, minimized with contrastive loss. The adaptive diagnostic head uses a Bayesian classifier to provide predictions of malignancy with uncertainty calibration and Grad-CAM++ to provide interpretability.

## IV. Results and Discussion

The Experimental analysis of the proposed HGGT framework indicates that it is effective in reliability and it is better than the traditional deep learning models in the diagnostic performance. The model is well balanced and strongly classified with accuracy, precision, recall and F1-score even in a case of class imbalance. The comparisons with HGGT show that it outperforms CNN, VGG16, ResNet50, EfficientNet-B0 and GAN-assisted CNN models on all metrics, which indicates that it has a good multimodal fusion capacity. The ablation study also confirms the fact that each individual aspect of the study including the attention mechanisms, CycleGAN augmentation as well as Swin-CNN-GAT integration contribute significantly to the performance increment. Grad-CAM++ visualizations reveal that the framework is effective at localizing malignant regions, as GAT attention maps show the importance of clinical biomarkers such as tumor size, ER as well as PR status. All in all, the results prove HGGT to be a fit, understandable, and valid model in the diagnosis of breast cancer.

TABLE I. Simulation Parameters

| Parameter | Value |
|---|---|
| Dataset | CBIS-DDSM |
| Image Resolution | 512 × 512 pixels |
| Methods used | Swin Transformer + CNN, Graph Attention Network, |
| Batch Size | 16 |
| Initial Learning Rate | 1e-4 with cosine annealing scheduler |
| Optimizer | AdamW |
| Epochs | 150 (with early stopping) |
| Dropout Rate | 0.3 |
| Weight Decay | 5e-5 |
| Cross-validation | 5-fold cross-validation |
| Evaluation Metrics | Accuracy, Sensitivity, Specificity, F1-score, AUC-ROC |
| Software | Python 3.10 |

Table I illustrates the HGGT framework simulation parameters are designed to give strong and yielding prediction of breast cancer. The data used in the analysis are CBIS-DDSM

in which the mammogram images have already been preprocessed to 512 512-pixels resolution. The model includes Swin Transformer + CNN with images and Graph Attention Network (GATT) with clinical data. AdamW optimizer is applied to train and initial learning rate is 1e-4 in the form of cosine annealing with a batch of 16 and 150 epochs total with early stopping. Regularization dropout rate 0.3 weight decay 5e-5. The 5-fold cross-validation was considered as evaluation methods and such metrics like accuracy, sensitivity, specificity, F1-score, and AUC-ROC were used.

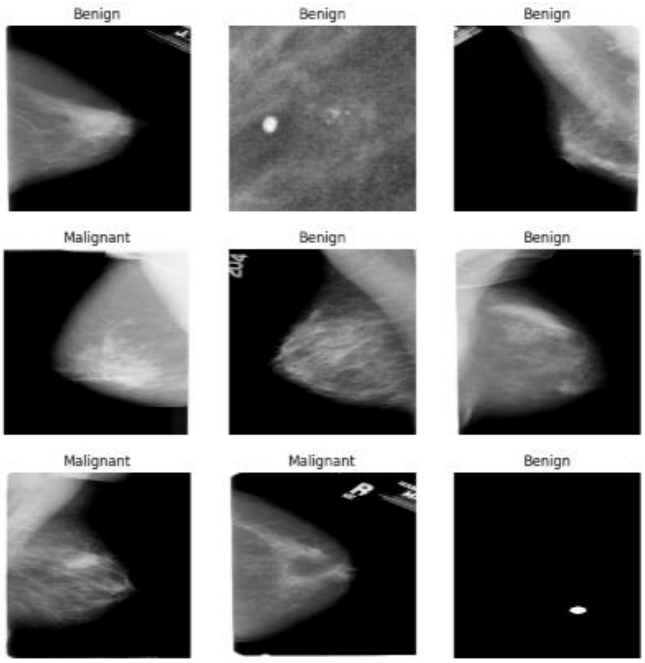


Fig 4. Breast image classification.

Fig. 4 shows a collection of mammogram images grouped as benign and malignant cases. The dataset grid has nine samples distributed across three rows and three columns, highlighting the variability in tissue patterns and tumor appearances. The benign images generally consist of localized or well-delineated areas with comparatively uniform density, while the malignant images exhibit irregularly dense structures with poorly defined borders, which are indicative of invasive growth.

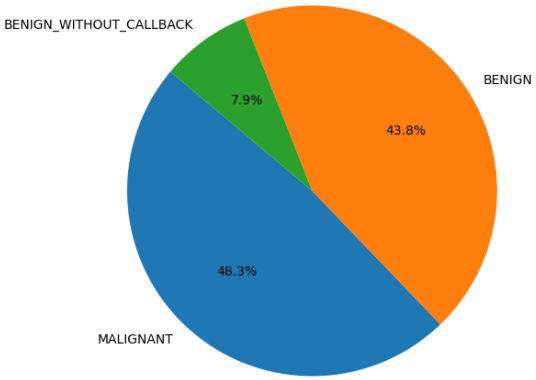


Fig 5. Class distribution.

Fig. 5 shows the division of the medical imaging results into three diagnostic types: Malignant, Benign, and Benign without Callback. Malignant conditions form the highest percentage at 48.3%, showing a high prevalence of confirmed malignancies. Benign conditions cover 43.8%, showing a significant percentage of non-cancerous findings. The lowest category, Benign without Callback, occupies merely 7.9%, showing the cases in which no follow-up was needed.

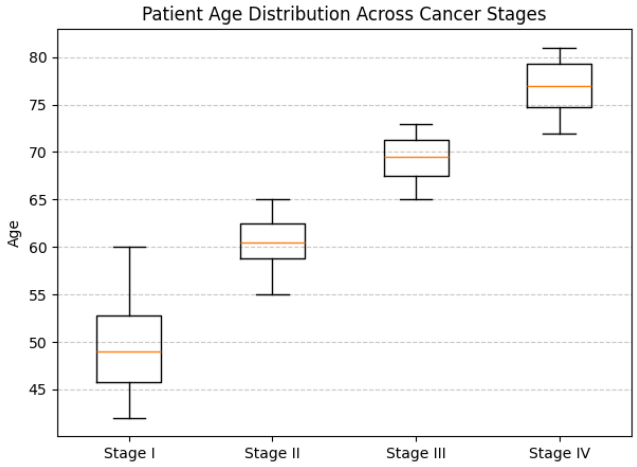


Fig 6. Patient age distribution across cancer stages.

Fig. 6 displays the distribution of the age of the patients in various stages of breast cancer. There is a separate box, denoted by each stage, Stage I, Stage II, Stage III, and Stage IV, which shows the median, interquartile range (IQR), and the minimum and maximum ages. In Stage I, the median age is less, at around 49 years, with a broader range of between 42 and 60 years, signifying that there is more variability in younger patients. The median ages are increasingly higher in Stage II and III at about 60 and 69 years, respectively, with smaller age ranges. The median age of patients in stage IV is the largest in 77 years, implying that later stages have more representation in older cohorts.

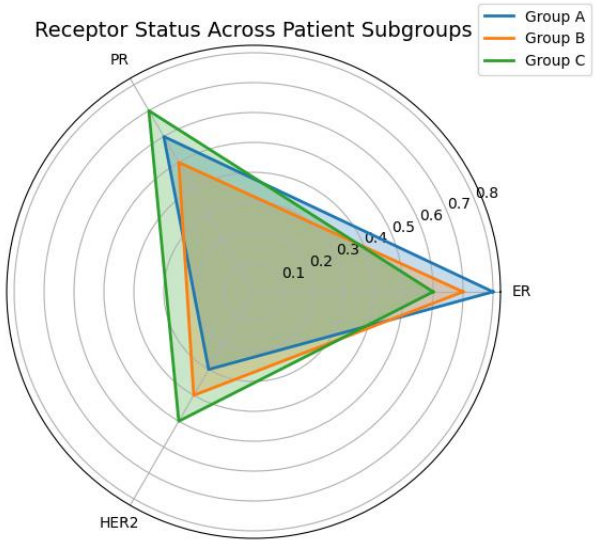


Fig 7. Receptor status across patient subgroups.

Fig. 7 depicts the receptor status (ER, PR, HER2) in three groups of patients in: Group A, Group B, and Group C. The axis depicts the value of each receptor type, and therometers are 0-0.8, which is the extent of receptor expression. Group A has the highest ER expression and Group C indicates the highest PR expression. HER2 expression is relatively low in all the groups, with a slight increment in Group C than Group A and B. The filled areas show differences in receptor patterns, with a particular emphasis on the fact that each subgroup has a different molecular profile.

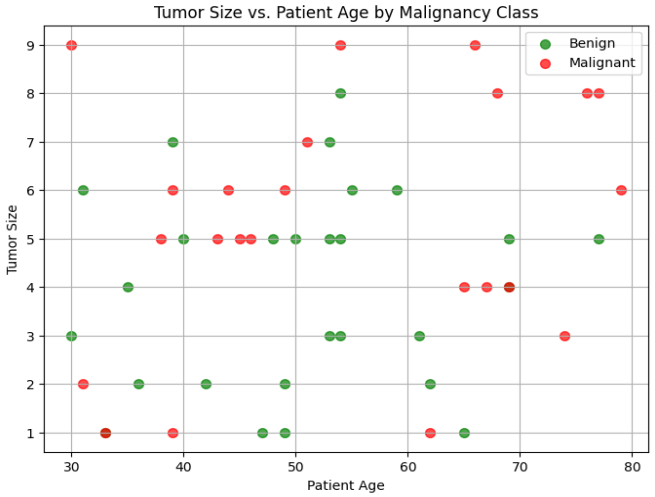


Fig 8. Tumor size vs. Patient age by malignancy class.

Fig. 8 can be used to show the correlation between the age of patients and the size of the tumor, according to the classification of malignancy. The points are the representatives of individual patients, and green markers are the indicator of benign tumors, whereas red markers denote the malignant tumors. The sizes of tumors are 1 to 9, whereas the ages of the patients do not exceed 30 to 80 years. The plot exposes that malignant tumors develop in a vast age bracket and are frequently larger in size, but there are also small malignant tumors. Benign tumor is evenly spread in terms of size and age.

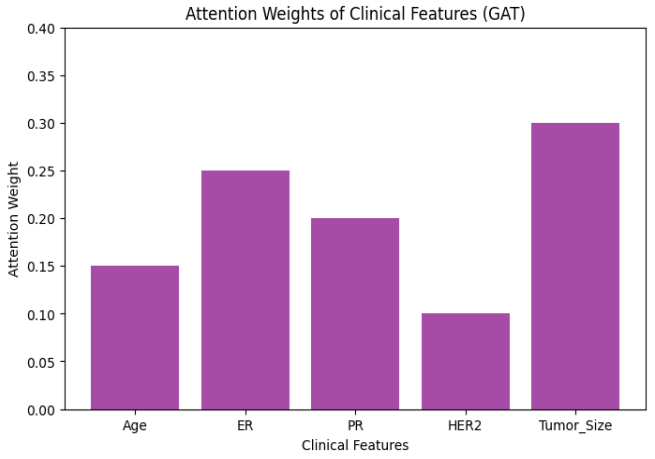


Fig 9. Attention weights of clinical features.

Fig. 9 depicts weights of attention of clinical features obtained using GAT model, and their significance in relation to their role in diagnosis prediction. Among the characteristics, the most important contribution is made by Tumor_Size, which suggests that it has a dominant contribution to clinical decision-making. ER and PR comes next with a medium level of importance as they play a great role in diagnostic interpretation. The smaller weight is supported by age, which implies a less direct but also relevant influence, whereas the contribution of HER2 is the lowest, indicating the lack of significance in this respect. In general, the image highlights that the size of tumor, and features of their receptors are the greatest clinical drivers that the model uses to make effective medical predictions.

### A. Performance Evaluation

Standard metrics are used to assess the categorization model's accuracy, while confusion matrix evaluation is used to measure false positives and misleading negatives.

*1) Accuracy*: Accuracy is a measure of the model's overall accuracy in prediction and is given in Eq. (11):

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn} \quad (11)$$

where, $tp$ is Rightly labeled malignant cases, $tn$ Rightly labeled benign cases, $fp$ is Wrongly labeled benign cases as malignant, $fn$: Wrongly labeled malignant cases as benign.

*2) Precision*: Precision calculates the ratio of malignant cases correctly identified out of all cases that are predicted to be malignant in Eq. (12):

$$Precision = \frac{tp}{tp+fp} \quad (12)$$

A high precision value means fewer false positives, and this is particularly important in medical diagnosis to prevent unnecessary biopsies.

*3) Recall*: Recall is the proportion of valid malignant cases that were properly identified in Eq. (13):

$$Recall = \frac{tp}{tp+fn} \quad (13)$$

It will have less false negatives and a high recall, which is important in breast cancer screening to limit missed diagnosis.

*4) F1-score*: The F1-score is a balance between Precision and Recall, hence it is more appropriate for imbalanced sets in Eq.(14):

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (14)$$

High F1-score indicates that the model is maintaining precision and recall in equilibrium.

TABLE II. PERFORMANCE METRICS

| Metrics | Value% |
|---|---|
| Accuracy | 98.2 |
| Precision | 98.7 |
| Recall | 98.5 |
| F1-Score | 99.2 |

Fig. 10 and Table II show the performance measures of the model, and it is effective in the four important measures. The

model has an accuracy of 98.2 % and demonstrates how well it performs on the predictions in general. The accuracy is 98.7 %, which shows that the model has a high capability of reducing false positives. The recall is also a bit lower (98.5 %), which indicates the ability of the model to report the important instances without recording the false negatives. Single accuracy of the F1-score, which combines both the precision and recall, has the highest value of 99.2 %, which is consistent and strong. Altogether, these measurements prove the great reliability of the model, harmonized predictive power, and its ability to be applicable in practice when performing medical diagnostic tasks.

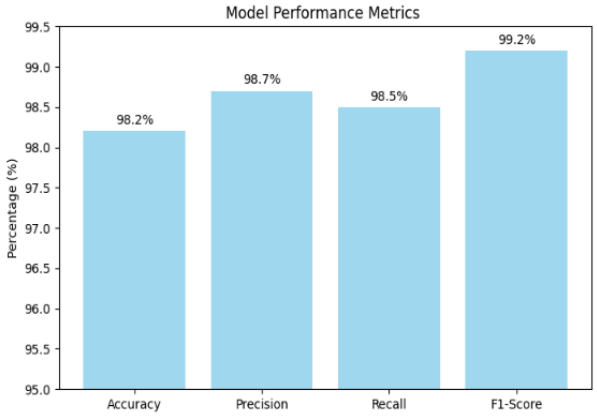


Fig 10. Performance metrics.

## B. Comparison Metrics

Table III and Fig. 11 introduce a comparison analysis of the performance of different deep learning models used in the diagnosis of breast cancer. The performance indicators like accuracy, precision, recall, F1 score, and AUC ROC of standard architectures, such as CNN, VGG16, ResNet50, and EfficientNet B0, show increasing improvement regarding all these metrics, which signifies improved feature extraction and classification performance. The hybrid of GAN and CNN is another way of improving the performance of predictive, and symbolizes the power of generative augmentation. It is worth noting that the proposed model has the best scores in accuracy, 98.2 %, precision, 98.5 %, recall, 99.2 % and F1 score, 99.1 % AUC ROC, indicating its superior diagnostic ability. These findings show that the hybrid method proposed is effective in detecting breast cancer using multimodal data.

TABLE III. PERFORMANCE COMPARISON

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) | AUC-ROC (%) |
|---|---|---|---|---|---|
| CNN [25] | 91.5 | 89.6 | 90.2 | 89.9 | 93.0 |
| VGG16 [26] | 94.2 | 93.5 | 94.0 | 93.7 | 95.6 |
| ResNet50 [27] | 95.0 | 94.1 | 95.5 | 94.8 | 96.5 |
| EfficientNet-B0 [28] | 95.6 | 95.0 | 95.7 | 95.3 | 97.2 |
| GAN + CNN [17] | 96.0 | 95.4 | 96.8 | 96.1 | 97.8 |
| Proposed Model | 98.2 | 98.7 | 98.5 | 99.2 | 99.1 |

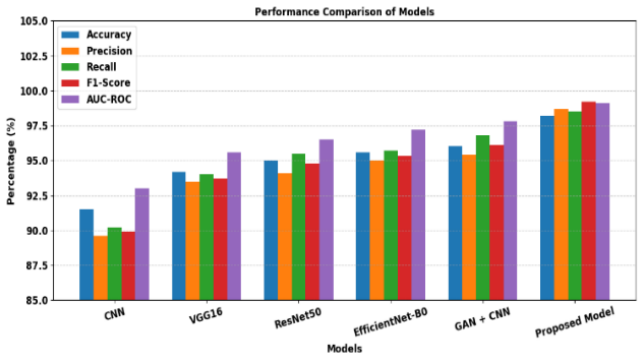


Fig 11. Performance metrics comparison.

TABLE IV. ABLATION STUDY

| Model Variant | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Swin Transformer Only | 94.8 | 94.0 | 95.1 | 94.5 |
| Swin + Attention Mechanisms | 96.1 | 95.5 | 96.8 | 96.1 |
| Swin + CycleGAN Synthetic Images | 96.8 | 96.2 | 97.5 | 96.8 |
| Swin Transformer-CNN + GAT | 98.2 | 98.7 | 98.5 | 99.2 |

Table IV reflects the performance comparison of various model variants that are developed on Swin Transformer in the diagnosis of breast cancer. The Swin Transformer with the basic architecture has a high baseline accuracy of 94.8 %. This addition of the mechanisms of attention enhances the model, leading to 96.1 % accuracy and an increase in precision, recall, and F1 score, which underscores the positive effects of accurate weighting of features. The use of CycleGAN-generated synthetic images further increases the accuracy to 96.8 % by increasing the training data. These results are the most significant when Swin Transformer-CNN is combined with GAT, as the accuracy is 98.2 %, the precision is 98.7 %, the recall is 98.5 %, and the F1 score is 99.2 %, which underlines the effectiveness of the multimodal fusion and further graph-based learning in enhancing the accuracy of the diagnosis.

## C. Discussion

The effectiveness of the HGGT framework can be explained by the fact that it is capable of modeling hierarchical mammographic representations and structured clinical relationships jointly, allowing to make more diagnostic reasoning than unimodal deep learning methods. The HGGT model is an important leap towards the multimodal detection of breast cancer because it is fast to integrate both mammographic images and clinical biomarkers into one diagnostic model. Unlike other traditional CNN-based networks, the HGGT is a hybrid Swin Transformer-CNN block that focuses on the visual characteristics to capture the global images of the breast and fine details of a lesion. Meanwhile, Graph Attention Network forecasts malignancy in biologic details by learning intricate clinical interactions between biomarkers, such as ER, PR, HER2, and tumor stage. The mechanism of contrastive pyramid fusion proves efficient in matching the imaging and clinical embeddings that increase the diagnostic consistency and interpretability. Local and regional fusion also serves to improve

the detection in the early stages by means of the linkage of quadrant-specific images to histopathological features. It has added a Bayesian diagnostic head to provide the estimates of uncertainties, which is an important feature of high-risk clinical decision-making. HGGT, with the help of interpretability agents of which Grad-CAM++ is a part, bridges the gap between AI predictions and clinician reasoning and constitutes a reliable, explainable, and clinically meaningful diagnostic model.

## V. Conclusion and Future Works

The research illustrates that successful diagnosis of breast cancer demands a coherent rationale of imaging and clinical modalities as opposed to solitary or superficial integration measures. It is demonstrated in the proposed Hybrid Graph-Generative Transformer (HGGT) that hierarchical multimodal fusion based on clinical relationships can enhance the reliability of diagnostic information, interpretability, and uncertainty awareness in computer-aided diagnosis systems. The framework, which integrates the representation of mammographies and clinically significant biomarker interactions, fills the gap between clinically relevant biomarker interactions and data-driven predictions, which help in making choices more transparent and reliable in obtaining trustful and clinician-focused decision support.

Although it has effective performances, this study has its limitations. It uses one retrospective dataset, and this might not be representative of other institutions and imaging platforms. Moreover, the framework can limit the real-time implementation to resource-constrained clinical environments because of its computational complexity. The areas of future work will be large-scale multi-institutional validation, computational optimization, and the combination of longitudinal, genomic, and radiomic data. Federated learning with privacy concerns and integration with clinical workflow are also major steps towards practical implementation.

## References

[1] S. Guan and M. Loew, "Using generative adversarial networks and transfer learning for breast cancer detection by convolutional neural networks," in Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications, SPIE, 2019, pp. 306–318.

[2] C. Singla, P. K. Sarangi, A. K. Sahoo, and P. K. Singh, "Deep learning enhancement on mammogram images for breast cancer detection," Mater. Today Proc., vol. 49, pp. 3098–3104, Jan. 2022, doi: 10.1016/j.matpr.2020.10.951.

[3] M. Behzadpour, B. L. Ortiz, E. Azizi, and K. Wu, "Breast tumor classification using efficientnet deep learning model," ArXiv Prepr. ArXiv241117870, 2024.

[4] Y. Jiménez-Gaona, D. Carrión-Figueroa, V. Lakshminarayanan, and M. J. Rodríguez-Álvarez, "Gan-based data augmentation to improve breast ultrasound and mammography mass classification," Biomed. Signal Process. Control, vol. 94, p. 106255, 2024.

[5] R. B. Eshun, A. K. Islam, and M. Bikdash, "A deep convolutional neural network for the classification of imbalanced breast cancer dataset," Healthc. Anal., p. 100330, 2024.

[6] M. S. A. Khan, A. Husen, S. Nisar, H. Ahmed, S. S. Muhammad, and S. Aftab, "Offloading the computational complexity of transfer learning with generic features," PeerJ Comput. Sci., vol. 10, p. e1938, 2024.

[7] Y. Jiménez-Gaona, M. J. Rodríguez-Alvarez, L. Escudero, C. Sandoval, and V. Lakshminarayanan, "Ultrasound breast images denoising using generative adversarial networks (GANs)," Intell. Data Anal., vol. 28, no. 6, pp. 1661–1678, 2024.

[8] R. Dijkstra, "The Effects of Data Augmentation and Synthetic Data in Breast Cancer Detection," 2024.

[9] L. L. Scientific, "LEVERAGING EXPLAINABLE AI TO IMPROVE BREAST CANCER DETECTION RATE USING TRANSFER LEARNING WITH DEEP RECURRENT CONVOLUTIONAL NEURAL NETWORKS," J. Theor. Appl. Inf. Technol., vol. 103, no. 1, 2025.

[10] J. Ellis, K. Appiah, E. Amankwaa-Frempong, and S. C. Kwok, "Classification of 2d ultrasound breast cancer images with deep learning," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2024, pp. 5167–5173.

[11] P. U. HEPSA, S. A. OZEL, and A. YAZICI, "Transfer Learning with Fuzzy for Breast Cancer," J. Inf. Sci. Eng., vol. 40, pp. 919–939, 2024.

[12] S. A. Qureshi et al., "Breast Cancer Detection Using Mammography: Image Processing to Deep Learning," IEEE Access, vol. 13, pp. 60776–60801, 2025, doi: 10.1109/ACCESS.2024.3523745.

[13] Y. E. Almalki, T. A. Soomro, M. Irfan, S. K. Alduraibi, and A. Ali, "Impact of Image Enhancement Module for Analysis of Mammogram Images for Diagnostics of Breast Cancer," Sensors, vol. 22, no. 5, Art. no. 5, Jan. 2022, doi: 10.3390/s22051868.

[14] Z. Jafari and E. Karami, "Breast Cancer Detection in Mammography Images: A CNN-Based Approach with Feature Selection," Information, vol. 14, no. 7, Art. no. 7, July 2023, doi: 10.3390/info14070410.

[15] S. Chaudhury and K. Sau, "Classification of Breast Masses Using Ultrasound Images by Approaching GAN, Transfer Learning, and Deep Learning Techniques," J. Artif. Intell. Technol., vol. 3, no. 4, pp. 142–153, 2023.

[16] W. H. Alawee, L. A. Al-Haddad, A. Basem, and A. A. Al-Haddad, "A data augmentation approach to enhance breast cancer detection using generative adversarial and artificial neural networks," Open Eng., vol. 14, no. 1, p. 20240052, 2024.

[17] E. Strelcenia and S. Prakoonwit, "Improving cancer detection classification performance using GANs in breast cancer data," IEEE Access, 2023.

[18] B. Swiderski, L. Gielata, P. Olszewski, S. Osowski, and M. Kołodziej, "Deep neural system for supporting tumor recognition of mammograms using modified GAN," Expert Syst. Appl., vol. 164, p. 113968, 2021.

[19] S. Sharma, Y. Singh, and T. Choudhury, "Advanced deep learning architectures for enhanced mammography classification: a comparative study of CNNs and ViT," Discov. Artif. Intell., vol. 5, no. 1, p. 187, 2025.

[20] S. Yang et al., "Mammo-Clustering: A Weakly Supervised Multi-view Global-Local Context Clustering Network for Detection and Classification in Mammography," ArXiv Prepr. ArXiv240914876, 2024.

[21] S. Fatima, M. U. Akram, S. Mohammad, and S. B. Ahmed, "Deep learning in dermatopathology: applications for skin disease diagnosis and classification," Discov. Appl. Sci., vol. 7, no. 9, pp. 1–26, 2025.

[22] M. Prodan, E. Paraschiv, and A. Stanciu, "Applying deep learning methods for mammography analysis and breast cancer detection," Appl. Sci., vol. 13, no. 7, p. 4272, 2023.

[23] Awsaf, "CBIS-DDSM: Breast Cancer Image Dataset," 2021, doi: 2021.

[24] "Real Breast Cancer Data." Accessed: Aug. 20, 2025. [Online]. Available: https://www.kaggle.com/datasets/amandam1/breastcancerdataset

[25] M. A. Yakoubi, N. Khiari, A. Khiari, and A. Melouah, "Deep Neural Network-Based Model for Breast Cancer Lesion Diagnosis in Mammography Images," Acta Inform. Pragensia, vol. 13, no. 2, pp. 213–233, 2024.

[26] A. Bechar, R. Medjoudj, Y. Elmir, Y. Himeur, and A. Amira, "Federated and transfer learning for cancer detection based on image analysis," Neural Comput. Appl., pp. 1–46, 2025.

[27] G. Ayana, K. Dese, A. M. Abagaro, K. C. Jeong, S.-D. Yoon, and S. Choe, "Multistage transfer learning for medical images," Artif. Intell. Rev., vol. 57, no. 9, p. 232, 2024.

[28] H. Avcı and J. Karakaya, "A novel medical image enhancement algorithm for breast cancer detection on mammography images using machine learning," Diagnostics, vol. 13, no. 3, p. 348, 2023.