MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3334-3339

PARALLEL DUAL-STREAM VISION TRANSFORMER NETWORK FOR BREAST CANCER CLASSIFICATION IN MAMMOGRAPHY IMAGES

Ms. Dhamayanthi P, Assistant Professor
Department of Computer Science and Engineering,
KGiSL Institute of Technology, Coimbatore, Tamil Nadu, India
dhamayanthi.p@kgkite.ac.in

Keerthi R V
Department of Computer Science and Engineering,
KGiSL Institute of Technology, Coimbatore, Tamil Nadu, India
keerthi.official.007@gmail.com

Mahesh kanna K
Department of Computer Science and Engineering,
KGiSL Institute of Technology, Coimbatore, Tamil Nadu, India
kmaheshkanna2005@gmail.com

Loganathan S
Department of Computer Science and Engineering,
KGiSL Institute of Technology, Coimbatore, Tamil Nadu, India
loganathan15025@gmail.com

Arunadevi D
Department of Computer Science and Engineering,
KGiSL Institute of Technology, Coimbatore, Tamil Nadu, India
arunadevidurairaj2@gmail.com

Abstract — One of the major health issues facing women globally is breast cancer, and the greatest strategy to increase patient survival rates and simplify treatment is still early identification. Because it allows doctors to detect worrisome abnormalities such masses, microcalcifications, and architectural distortions inside breast tissue, mammography is regarded as the gold standard imaging technique for breast cancer screening. However, a variety of factors, such as the subtle appearance of early-stage tumors, differences in breast density, and the existence of structural structures that may visually resemble malignant lesions, make it difficult to appropriately interpret mammograms. Sometimes, even seasoned radiologists fail to notice minor irregularities or mistake benign patterns for malignant resultsBecause of these challenges, a lot of research has been done on computer-aided diagnostic systems that can assist radiologists by providing automated analysis and decision support.In recent years, deep learning has demonstrated remarkable achievements in medical image processing, particularly for radiology classification and detection applications. Convolutional neural networks have long dominated this discipline because of their ability to extract hierarchical visual information from images. However, because CNN-based models often rely on local receptive fields, they may have trouble capturing long-range spatial correlations across huge medical images, such as mammograms.More importantly, several studies have shown that deep learning models can be impacted by shortcut learning, where the network learns to rely on non-medical cues like image borders, scanner artifacts, or high-contrast anatomical structures rather than the actual pathological features associated with disease. This habit may lead to low reliability but deceptively great accuracy during training when applied in real clinical settings. Thus, there is an increasing need for more robust architectures that can capture both global anatomical context and fine-grained disease features while avoiding bias toward irrelevant visual patterns.

I.INTRODUCTION

One of the most prevalent cancers impacting women globally is breast cancer. Millions of new cases of breast cancer are discovered year, according to international health organizations, and early detection is critical to increasing treatment outcomes and survival rates. Early detection greatly improves the likelihood of a successful course of therapy for breast cancer. Because of this, hospitals and diagnostic facilities frequently employ screening techniques like mammography to find anomalies in breast tissue before the disease reaches an advanced stage.Mammography is thought to be the best imaging method for screening for breast cancer since it may identify tissue deformities, calcifications, and tiny lumps that might be signs of malignancies. But it's not always simple to analyze mammography pictures. Tumors can have extremely modest appearances, and some benign formations may resemble malignant lesions. Interpretation may also be more difficult due to elements including thick breast tissue, picture noise, and differences in scanning circumstances. Even skilled radiologists occasionally overlook minor anomalies or misclassify particular patterns, which could result in a delayed diagnosis or needless medical procedures. Researchers have created computer-aided diagnosis (CAD) systems that employ artificial intelligence to identify and categorize anomalies in medical pictures in order to help radiologists analyze mammograms more effectively. Deep learning methods have greatly enhanced these systems' performance in the last few years. Because convolutional neural networks (CNNs) can automatically extract key visual properties from images without the need for manual feature extraction, they have found widespread application in medical image analysis. Despite their effectiveness, conventional CNN models may have trouble capturing long-range correlations across large pictures, such as mammograms, because they primarily concentrate on local image elements. Vision Transformers (ViTs) have become effective models for image analysis jobs in more recent times. Transformer models, in contrast to CNNs, employ self-attention processes that enable them to concurrently examine correlations between various aspects of a picture. They are better able to capture local and global information thanks to this capability. However, because mammograms contain both very fine specific patterns that may indicate disease and large-scale structural information, using a single model may still limit the system's ability to completely comprehend complicated medical pictures. To increase model resilience, the suggested system uses a variety of training methodologies and data augmentation techniques in addition to the architecture design. The objective of this research is to create a dependable and comprehensible deep learning system that can help radiologists detect breast cancer more accurately. The suggested method aims to replicate the diagnostic reasoning process employed by medical professionals by integrating global and local feature extraction into a single framework. These systems could facilitate clinical decision-making, lessen the burden of diagnosis, and eventually aid in the early detection of breast cancer.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3334-3339

II. LITERATURE SURVEY

For many years, research on the use of medical imaging for breast cancer detection has been ongoing. Because it enables medical professionals to discover anomalies in breast tissue before to the onset of symptoms, mammography is largely regarded as the most dependable screening technique for early identification of breast cancer. However, because of the intricate structure of breast tissue, variations in picture quality, and the existence of thick tissue that may conceal cancers, interpreting mammograms is frequently difficult. Researchers have created computer-aided diagnosis systems that automatically evaluate mammography images and identify worrisome areas using machine learning and deep learning approaches in order to assist radiologists and lower diagnostic errors. Traditional machine learning techniques that depended on manually created characteristics were the main focus of early research on breast cancer detection. In order to determine if a tumor was benign or malignant, researchers employed classification algorithms like Support Vector Machines, Decision Trees, or Random Forests after extracting picture attributes including texture, shape, and intensity using image processing techniques. These methods produced encouraging outcomes, although they mostly relied on expertly built features. Deep domain expertise was needed to design these features, and the models' effectiveness frequently hinged on how well these features represented the traits of tumors. Convolutional neural networks (CNNs), which can automatically learn hierarchical features directly from raw images, have become widely employed in medical image analysis with the development of deep learning.

VGGNet, ResNet, and DenseNet are examples of CNN-based architectures that have been effectively used for breast cancer screening applications. Important visual patterns such edges, textures, and shapes that are connected to aberrant tissue can be recognized by these networks. CNN-based algorithms can categorize mammograms into benign and malignant categories with excellent accuracy, as several studies have shown. Long-range spatial correlations within large medical images, such as mammograms, may be difficult for CNN models to grasp since they tend to concentrate on local image elements. In order to enhance the effectiveness of breast cancer detection models, researchers have also looked into transfer learning. Transfer learning is the process of fine-tuning deep neural networks on medical datasets after they have been pretrained on massive image datasets like ImageNet. This method uses information gained from general picture recognition tasks to help overcome the problem of limited medical data. Numerous studies have demonstrated that, in comparison to models trained from scratch, transfer learning greatly enhances classification performance. Because medical imaging patterns are different from common objects found in natural images, transfer learning from natural image datasets may not always capture the distinctive features of medical images, despite its benefits. Vision Transformers have become a potent substitute for convolutional neural networks in image classification problems in more recent times. Vision Transformers examine connections between various areas of a picture by using self-attention mechanisms. By enabling each image patch to interact with every other patch, transformers are able to collect global contextual information in contrast to CNNs, which process images via localized filters. Because of this feature, transformers are especially helpful for processing complicated images where contextual linkages and global structure are crucial.

S.No	Title	Author	Year
1	Deep Learning Algorithms for Breast Cancer Detection	McKinney et al.	2024
2	XGBoost for Breast Cancer Classification	Islam et al.	2024
3	Vision Transformer Transfer Learning for Mammogram Classification	Shen et al.	2023
4	Ensemble Methods for Medical Image Classification	Mahesh et al.	2022
5	Interpretable Machine Learning for Breast Cancer Diagnosis	Dutta et al.	2024

Table 1.1 Literature Survey

When compared to conventional CNN-based models, Vision Transformers have been shown to perform better in a number of studies that have used them for medical imaging applications, such as breast cancer diagnosis. Because of its effective attention mechanism and hierarchical design, the Swin Transformer has drawn a lot of attention among transformer-based systems. Shifted window attention is used by the Swin Transformer to interpret images, allowing the model to retain global context while capturing local details. Tumor detection and segmentation are just two of the medical image processing tasks in which this architecture has proven effective. It can detect tiny abnormalities in mammography images, like microcalcifications and aberrant tissue structures, because of its capacity to evaluate fine-grained texture patterns.

III. PROBLEM STATEMENT

One of the main causes of death for women globally is breast cancer, and improving treatment results and survival rates depends heavily on early identification. Because mammography can identify early indicators of malignancies before physical symptoms manifest, it is frequently employed as the main screening method for identifying abnormalities in breast tissue. Mammography image interpretation, however, is a challenging and time-consuming task. Breast cancer can cause subtle visual patterns that resemble normal tissue structures, such as lumps, microcalcifications, and architectural aberrations. Because of this, even skilled radiologists can occasionally overlook early-stage cancers or mistakenly identify benign structures as malignant discoveries. The automatic identification of breast cancer from mammography pictures has been greatly aided by deep learning models in recent years. Convolutional Neural Networks (CNNs) and other deep learning architectures have demonstrated encouraging outcomes in the classification of medical pictures and the extraction of visual information. Despite these developments, there are still a number of obstacles. Local texture patterns or global structural information are the primary emphasis of many current models, but not both at the same time. The overall anatomy of the breast, the distribution of tissue density, and minute irregularities like microcalcifications are only a few of the many scales of complicated information found in mammography images. While models that solely concentrate on global structure could miss small lesions, models that solely concentrate on local traits might overlook crucial contextual information. The issue of shortcut learning is a significant obstacle in automated mammography analysis. Sometimes, instead of the real tumor patterns inside glandular tissue, deep learning models develop to rely on non-medical cues like high-contrast breast boundaries, pectoral muscle regions, or imaging abnormalities. In recent years, deep learning models have made a significant contribution to the autonomous detection of breast cancer using mammography images. Convolutional Neural Networks (CNNs) and other deep learning architectures have shown promising results in the extraction of visual information and the classification of medical images. There are still many challenges in spite of these advancements. Many existing models focus on either global structural information or local texture patterns, but not both at the same time. Mammography scans provide a wide range of complex information, including the distribution of tissue density, the overall morphology of the breast, and small anomalies like microcalcifications.

Models that focus just on local characteristics may lose important contextual information, whereas models that focus only on global structure may miss minor lesions. muscle. This behavior lowers the system's dependability in actual clinical settings and may result in inaccurate predictions. In order to direct the model to concentrate on areas of the image that are medically relevant, more

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3334-3339

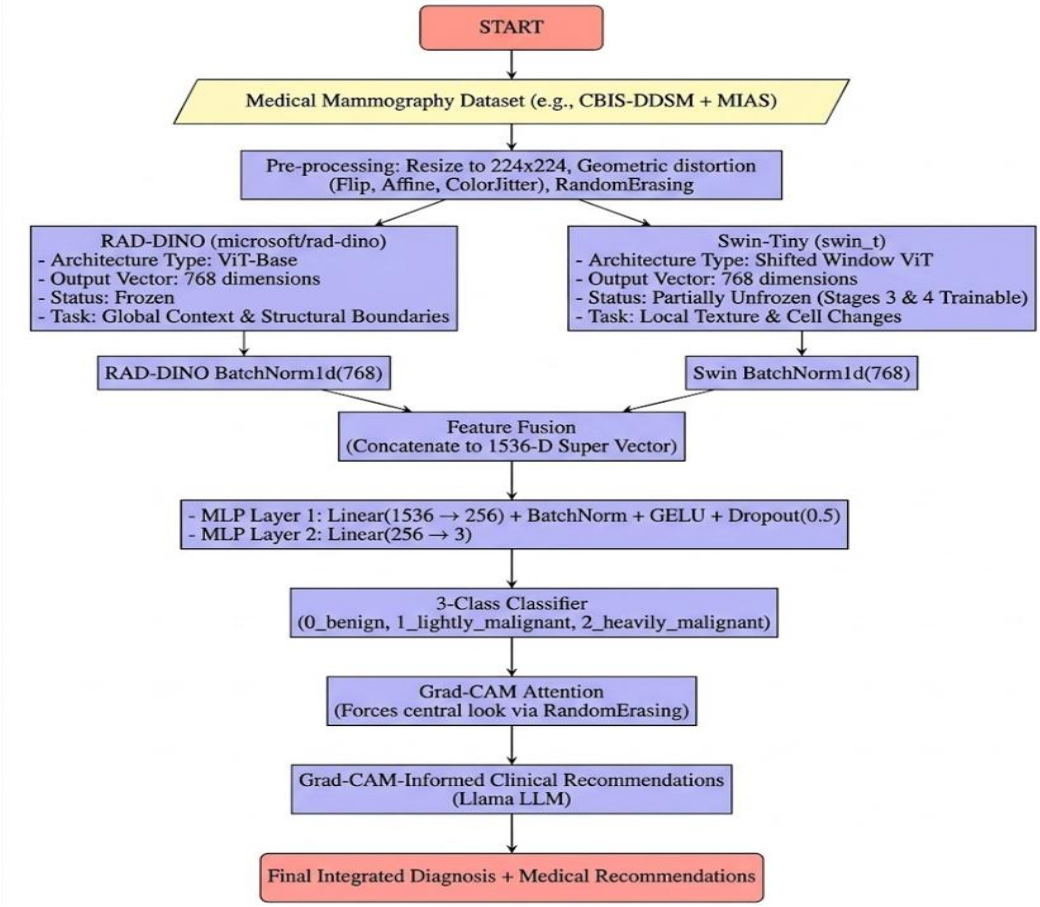
reliable and clinically significant architectures are required. Additionally, the size of medical imaging datasets is frequently smaller than that of natural picture datasets used in conventional computer vision tasks. It is challenging to efficiently train large deep learning models without overfitting due to this constraint. Even though this problem has been addressed with transfer learning and pretrained models, many current methods still have trouble generalizing well across various datasets and imaging settings. In light of these difficulties, the primary issue this study attempts to solve is how to create an efficient deep learning architecture that can minimize reliance on superfluous visual cues while simultaneously capturing global anatomical structure and fine-grained local texture information from mammography images .By integrating structural knowledge with in-depth analysis of questionable areas, the system should be able to evaluate breast images in a manner similar to radiologists' diagnostic procedures. This work suggests a dual-stream deep learning architecture that combines two complementary transformer-based models in order to address this issue. A pretrained Vision Transformer tailored to radiology is used in the first stream to extract global contextual information, and a hierarchical transformer architecture is used in the second stream to capture local texture patterns. The suggested method seeks to create a more complete representation of mammography pictures and enhance the precision and dependability of breast cancer classification by merging the data taken from both streams using a feature fusion technique. Therefore, the goal of this research is to create a reliable and understandable computer-aided diagnostic system that may help radiologists identify breast cancer more precisely and effectively, thereby promoting earlier diagnosis and improved patient outcomes.

IV. PROPOSED SOLUTION

To address the limitations of existing breast cancer detection systems, this work proposes a dual-stream transformer-based architecture that combines global anatomical understanding with detailed local texture analysis. The key objective of the proposed system is to mimic the diagnostic reasoning process followed by radiologists when analyzing mammography images.In clinical practice, radiologists first examine the overall structure of the breast and then carefully analyze suspicious regions for subtle abnormalities such as microcalcifications or irregular tumor margins. Inspired by this workflow, the proposed framework integrates two complementary Vision Transformer models that operate in parallel and extract different types of features from the same mammogram image. The architecture consists of two major feature extraction streams followed by a feature fusion module and a classification head. The first stream captures global structural information using a radiology-specific pretrained transformer, while the second stream focuses on identifying local texture patterns that may indicate the presence of malignant tumors. The outputs from these streams are combined to produce a comprehensive representation of the mammogram image, which is then used for multi-class classification.

A. Global Context Feature Extraction

The first stream of the proposed architecture focuses on extracting global contextual information from mammogram images. This branch employs the RAD-DINO Vision Transformer model, which has been pretrained using the DINO self-supervised learning framework on a large-scale radiological image dataset. Because the model has already learned meaningful representations of anatomical structures from extensive radiological data, it is particularly well suited for analyzing medical images.



In this stage, the input mammogram image is divided into small patches that are processed through multiple transformer layers using self-attention mechanisms. The self-attention operation allows each image patch to interact with all other patches, enabling the network to capture long-range dependencies across the entire image. This capability is important for mammography analysis because abnormalities may appear in relation to surrounding tissue structures rather than as isolated features.To preserve the pretrained knowledge of RAD-DINO, all parameters of the model remain frozen during the training phase. This prevents the network from overfitting to the limited mammography dataset while maintaining its ability to extract high-quality radiological features. The output of this stream is a 768-dimensional feature vector representing the global anatomical context of the breast image.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3334-3339

B. Local Texture Feature Extraction

While global structural information provides important contextual understanding, detecting breast cancer also requires detailed analysis of small tissue patterns. Malignant tumors often appear as subtle variations in tissue texture, such as clusters of microcalcifications, irregular margins, or localized distortions in glandular tissue.

To capture these fine-grained details, the second stream of the architecture utilizes the Swin Transformer model. The Swin Transformer is a hierarchical vision transformer that uses shifted window attention mechanisms to efficiently process local regions of the image. Instead of performing global attention across the entire image, the model divides the image into smaller windows and computes attention within each window.In the proposed system, the early layers of the Swin Transformer remain frozen to preserve general visual feature extraction capabilities, while the deeper layers are fine-tuned during training to adapt to the mammography domain. Similar to the RAD-DINO branch, the Swin Transformer produces a 768-dimensional feature vector representing local texture features.

C. Feature Fusion Mechanism

After both streams generate their respective feature representations, the next step is to combine them into a unified feature space. Since the RAD-DINO and Swin Transformer networks produce features with different statistical distributions, directly concatenating them may lead to unstable training.To address this issue, batch normalization layers are applied to both feature vectors before fusion. Once normalized, the two 768-dimensional vectors are concatenated into a 1536-dimensional fused feature vector. This fused representation integrates both global structural information and detailed local texture patterns extracted from the mammogram image

D. Classification Module

The fused feature vector is passed into a multi-layer perceptron (MLP) classification module. The first fully connected layer reduces the dimensionality from 1536 to 256 units, allowing the network to focus on the most informative features.Batch normalization and the GELU activation function are applied to improve model stability. In addition, dropout with a probability of 0.5 is used to reduce overfitting.

The final output layer maps the processed representation to three diagnostic classes:

- Benign
- Lightly Malignant
- Heavily Malignant

A softmax function converts the outputs into class probabilities.

E. Training and Optimization

The proposed architecture is trained using the AdamW optimizer, which effectively handles weight decay in transformer models. The initial learning rate is set to 3×10^{-4}, and a cosine annealing scheduler gradually decreases the learning rate during training.The model is trained for 35 epochs with an effective batch size of 16. Because GPU memory limits batch size per iteration, gradient accumulation is used to simulate larger batch sizes. Training is optimized using cross-entropy loss, which measures the difference between predicted class probabilities and ground-truth labels.Overall, the proposed dual-stream transformer architecture integrates both global and local feature extraction mechanisms to provide a comprehensive analysis of mammography images, thereby improving the accuracy and reliability of automated breast cancer classification.

V. SYSTEM ARCHITECTURE

The proposed system architecture is designed to improve the reliability and accuracy of automated breast cancer detection using mammography images. The overall framework follows a dual-stream deep learning approach in which two transformer-based models analyze the same input image simultaneously. Each stream focuses on extracting different types of information from the mammogram image.One branch captures the global anatomical context of the breast, while the second branch extracts detailed local texture features that may indicate the presence of tumors. By combining these complementary representations, the system attempts to replicate the diagnostic reasoning process used by radiologists when interpreting mammograms.

A. Input Pre processing

Before feeding mammogram images into the neural network, a pre processing stage is applied to ensure consistency and improve training performance. The input images obtained from the dataset are first resized to a fixed resolution of 224 × 224 pixels. Standardizing the input resolution allows the transformer networks to process images efficiently and ensures compatibility with pretrained models.In addition to resizing, the pixel intensity values of the images are normalized so that the distribution of pixel values remains consistent across the dataset. Normalization helps stabilize the training process and improves convergence during optimization.

Data augmentation techniques are also applied during training in order to increase dataset diversity and reduce overfitting. These transformations include horizontal flipping, small geometric rotations, brightness adjustments, and random erasing. Such augmentations simulate variations that may occur during real mammography scans and encourage the model to learn robust feature representations.

B. Global Context Extraction using RAD-DINO

The first branch of the architecture is responsible for capturing the global structural information present in the mammogram image. This stream uses the RAD-DINO Vision Transformer model, which has been pretrained on a large collection of radiological images using the DINO self-supervised learning framework. The RAD-DINO model divides the input image into smaller patches and processes them through multiple transformer layers using self-attention mechanisms.

Through this attention mechanism, each image patch can interact with every other patch in the image, enabling the network to capture long-range spatial relationships. As a result, the model can learn important anatomical features such as breast shape, tissue distribution, and density patterns. These global features provide valuable contextual information that assists in identifying abnormal tissue structures.

Since RAD-DINO already possesses strong radiological feature extraction capability, its parameters are kept frozen during training. Freezing the pretrained weights prevents the network from overfitting to the limited mammography dataset while preserving its ability to extract meaningful anatomical representations. The output of this branch is a 768-dimensional feature vector representing the global context of the mammogram image.

C. Local Texture Extraction using Swin Transformer

While global context is important, accurate breast cancer detection also requires detailed analysis of local tissue patterns. Malignant tumors often appear as subtle visual patterns such as microcalcifications, irregular margins, or small distortions in breast tissue. These patterns may occupy only a small portion of the image and therefore require fine-grained feature extraction.To capture such localized features, the second stream of the architecture employs the Swin Transformer model. The Swin Transformer is a hierarchical Vision Transformer that processes images using shifted window attention mechanisms. Instead of applying attention across the entire image, the model divides the image into smaller windows and performs self-attention operations within each window.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3334-3339

This design allows the network to efficiently capture detailed local features while maintaining computational efficiency. As the windows shift across layers, the model gradually integrates information from neighboring regions, enabling it to understand both local and regional structures. In the proposed architecture, the early layers of the Swin Transformer remain frozen while the deeper layers are fine-tuned during training. This strategy allows the network to retain general visual knowledge while adapting higher-level features specifically for mammography analysis. Similar to the first branch, this stream also produces a 768-dimensional feature vector representing the local texture characteristics of the mammogram.

D. Feature Fusion

After both feature extraction streams generate their respective representations, the architecture combines them through a feature fusion mechanism. Because the feature vectors produced by RAD-DINO and Swin Transformer originate from different network architectures, their distributions may vary significantly. To address this issue, batch normalization layers are applied to both feature vectors before fusion.Batch normalization standardizes the scale and distribution of the features, ensuring that both streams contribute equally to the final representation. Once normalized, the two 768-dimensional feature vectors are concatenated to form a fused feature vector of dimension 1536. This fused representation integrates both global anatomical context and fine-grained local texture information, providing a comprehensive description of the mammogram image.

E. Classification Head

The fused feature representation is then passed into a classification module implemented as a multi-layer perceptron (MLP). The first fully connected layer reduces the dimensionality of the fused feature vector from 1536 to 256 units. This dimensionality reduction helps the model focus on the most informative features extracted from the previous stages.To improve training stability and generalization performance, the output of the dense layer is processed through batch normalization and the Gaussian Error Linear Unit (GELU) activation function. Dropout regularization with a probability of 0.5 is also applied to prevent overfitting by randomly disabling neurons during training.

Finally, the processed features are passed to the output layer, which produces three probability values corresponding to the diagnostic classes: benign, lightly malignant, and heavily malignant. The Softmax function converts the outputs into class probabilities, and the class with the highest probability is selected as the final prediction.

F. Training Strategy

The proposed architecture is trained using the AdamW optimizer, which is well suited for transformer-based networks. AdamW separates weight decay from gradient updates, helping prevent overfitting and improving convergence. The initial learning rate is set to 3×10^{-4}, and a cosine annealing learning rate scheduler is used to gradually reduce the learning rate during training.Training is performed for 35 epochs with an effective batch size of 16. Because of GPU memory limitations, gradient accumulation is used to simulate larger batch sizes. The loss function used for optimization is cross-entropy loss, which measures the difference between predicted probabilities and ground-truth class labels.

Overall, the proposed dual-stream architecture enables the model to simultaneously analyze both global and local information present in mammogram images. By combining transformer-based feature extraction with a robust fusion strategy, the system provides a comprehensive framework for automated breast cancer classification.

VI. EXPECTED OUTCOMES

The proposed dual-stream transformer-based framework is expected to significantly improve the performance and reliability of automated breast cancer detection using mammography images. By combining the strengths of two complementary Vision Transformer architectures, namely **RAD-DINO** and **Swin Transformer**, the system aims to capture both global anatomical context and fine-grained local texture patterns present in mammogram images. This hybrid approach is expected to provide a more comprehensive understanding of breast tissue structures compared to traditional single-model architectures.

One of the primary expected outcomes of the proposed system is improved classification accuracy in identifying different types of breast abnormalities. The model is designed to classify mammogram images into three clinically meaningful categories: **benign, lightly malignant, and heavily malignant**. By incorporating both global and local feature representations, the system is expected to reduce misclassification errors that may occur when a model focuses only on one type of visual information. In particular, the architecture should demonstrate improved sensitivity in detecting malignant cases, which is essential for early cancer diagnosis.

Another important expected outcome is improved model interpretability and clinical reliability. Many conventional deep learning models behave as **black-box systems**, making it difficult for clinicians to understand the reasoning behind their predictions. The proposed framework integrates **attention-based feature extraction mechanisms** that highlight important regions in the mammogram image during the classification process. These visual explanations can help radiologists verify whether the model is focusing on relevant anatomical structures, thereby increasing trust in the automated system.

The proposed approach is also expected to demonstrate robustness across diverse mammography datasets. Through the use of pretrained transformer models and extensive **data augmentation techniques**, the architecture should be capable of learning generalized representations of breast tissue structures. This capability may allow the system to perform consistently across images obtained from different mammography devices, scanning conditions, and patient populations.Furthermore, the fusion of features from the **RAD-DINO** and **Swin Transformer** models is expected to reduce the risk of shortcut learning. Instead of relying on non-medical cues such as image borders or pectoral muscle regions, the model will be guided to focus on clinically relevant areas of glandular tissue. This improvement in feature learning can help produce more reliable predictions in real-world clinical settings.

Overall, the expected outcome of this research is the development of an effective **computer-aided diagnostic system** that can assist radiologists in identifying breast cancer more accurately and efficiently. By improving detection accuracy, interpretability, and robustness, the proposed system has the potential to support early diagnosis and ultimately contribute to better patient outcomes in breast cancer screening programs.

VII.CONCLUSION

Breast cancer remains one of the leading causes of mortality among women worldwide, and **early detection is crucial** for improving survival rates and treatment outcomes. Mammography screening plays a vital role in detecting abnormalities in breast tissue; however, interpreting mammogram images is a challenging and time-consuming task even for experienced radiologists. The complexity of breast tissue structures, variations in image quality, and the subtle appearance of tumors can lead to diagnostic errors or missed detections. These challenges highlight the need for reliable **computer-aided diagnostic** The proposed architecture demonstrates how **transformer-based models** can be effectively applied to medical image analysis tasks. Unlike traditional convolutional neural networks that primarily focus on local features, transformer architectures are capable of capturing **long-range dependencies and contextual relationships** within images. This capability allows the model to analyze complex breast tissue structures more effectively and detect subtle abnormalities that might otherwise be overlooked.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3334-3339

Another important contribution of this work is the use of **feature fusion** to integrate information from two different transformer architectures. The fusion mechanism combines global and local features to create a richer representation of the input image, which helps the classifier make more accurate predictions. This approach mimics the diagnostic workflow used by radiologists, where both overall anatomical structure and detailed tissue patterns are considered when making a diagnosis.Although the proposed system shows promising potential for improving automated breast cancer detection, there are still opportunities for further research and improvement. Future work may explore the integration of additional imaging modalities such as **ultrasound or MRI** to provide complementary diagnostic information. Furthermore, larger and more diverse datasets could be used to further enhance the generalization capability of the model.

In conclusion, the proposed dual-stream transformer architecture represents a promising approach for automated breast cancer detection in mammography images. By combining global contextual analysis with fine-grained texture recognition, the system offers improved **accuracy, interpretability, and robustness** compared to conventional single-model approaches. With further validation and development, such **AI-based diagnostic tools** may play an important role in supporting radiologists and improving early detection of breast cancer in clinical practice.

VIII.REFERENCES

[1] World Health Organization, *"Breast Cancer Fact Sheet,"* 2023.
[2] American Cancer Society, *"Breast Cancer Statistics,"* **CA: A Cancer Journal for Clinicians**, 2023.
[3] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. Aerts, *"Artificial intelligence in radiology,"* **Nature Reviews Cancer**, 2018.
[4] G. Litjens et al., *"A survey on deep learning in medical image analysis,"* **Medical Image Analysis**, 2017.
[5] R. Lee et al., *"The CBIS-DDSM dataset for mammography,"* **Scientific Data**, 2017.
[6] J. Suckling et al., *"The Mammographic Image Analysis Society (MIAS) database,"* 1994.
[7] R. Ragab et al., *"Transfer learning for medical imaging,"* **MICCAI**, 2023.
[8] Z. Liu et al., *"Swin Transformer: Hierarchical vision transformer using shifted windows,"* **ICCV**, 2021.
[9] Z. Liu et al., *"A ConvNet for the 2020s,"* **CVPR**, 2022.
[10] R. Selvaraju et al., *"Grad-CAM: Visual explanations from deep networks,"* **International Journal of Computer Vision**, 2020.
[11] F. Wu et al., *"CNN-based mammography classification using DenseNet,"* **Diagnostics**, 2023.
[12] M. Dutta et al., *"Interpretable machine learning for breast cancer diagnosis,"* **Bulletin of Electrical Engineering and Informatics**, 2024.
[13] E. Topol, *"High-performance medicine: The convergence of human and artificial intelligence,"* **Nature Medicine**, 2019.
[14] K. He, X. Zhang, S. Ren, and J. Sun, *"Deep residual learning for image recognition,"* **CVPR**, 2016.
[15] A. Dosovitskiy et al., *"An image is worth 16×16 words: Transformers for image recognition at scale,"* **ICLR**, 2021.
[16] T. Mahesh et al., *"Performance analysis of XGBoost ensemble methods,"* **Healthcare**, 2022.
[17] F. Liu et al., *"Multi-modal breast cancer detection,"* **Medical Image Analysis**, 2024.
[18] M. Rahman et al., *"A comprehensive survey on deep learning in breast cancer detection,"* **IEEE Access**, 2023.
[19] H. Shin et al., *"Deep learning in mammography: Current status and future perspectives,"* **Korean Journal of Radiology**, 2024.
[20] A. Hamidinekoo et al., *"Deep learning for mammogram analysis,"* **Physics in Medicine and Biology**, 2021.
[21] L. Shen et al., *"Deep active learning for breast cancer segmentation,"* **Medical Image Analysis**, 2021.
[22] Y. Zhang et al., *"Attention networks for medical image classification,"* **IEEE Transactions on Medical Imaging**, 2023.
[23] R. Geirhos et al., *"Shortcut learning in deep neural networks,"* **Nature Machine Intelligence**, 2020.
[24] S. Sagadeeva et al., *"Shortcut learning leads to bias in medical diagnosis,"* **medRxiv**, 2024.
[25] A. Esteva et al., *"Dermatologist-level classification of skin cancer with deep neural networks,"* **Nature**, 2017.
[26] A. Madani et al., *"Fast and accurate view classification of mammograms,"* **IEEE ISBI**, 2018.
[27] J. Ker et al., *"Deep learning applications in medical image analysis,"* **Journal of X-Ray Science and Technology**, 2018.
[28] S. Anwar et al., *"Medical image analysis using convolutional neural networks,"* **Journal of Medical Systems**, 2018.
[29] P. Tang et al., *"Automated breast cancer detection in mammography using deep learning,"* **IEEE Transactions on Medical Imaging**, 2023.
[30] Hugging Face, *"LLaMA: Open foundation language models,"* **Meta AI**, 2023.