MetaBreastAI: An Explainable Dual-Stream Convolutional Neural Network-Transformer Framework with Multi-Instance Learning for Breast Cancer Metastasis Detection

Bnar M. Ghafour[†] and Abbas M. Ali

[1]Department of Software Engineering and Informatics, College of Engineering, Salahaddin University-Erbil, Kurdistan Region – F.R. Iraq

***Abstract*–Breast cancer metastasis is one of the most consequential clinical problems in breast cancer, as it is responsible for most of the breast cancer attributable deaths in women globally. Timely detection of metastasis progression is crucial for improving therapies and enhancing patient survival. Although deep learning approaches, especially convolutional neural networks (CNNs), still fail to model global dependencies, they cannot work under weak supervision, and unable to generate interpretable predictions. Transformer-based models, on the other hand, provide a deeper contextual knowledge; but they are insensitive to local patterns and require large datasets. To overcome these issues, we present MetaBreastAI, an explainable dual-stream deep learning framework comprising a CNN branch fused with the convolutional block attention module and a Transformer stream. Each of these parallel branches captures spatial and contextual features, respectively, which are fused by a multi-instance learning approach for partially supervised classification. The proposed model has been tested on benchmark computed tomography scan dataset, it enables identifying one of the distinct features of MetaBreastAI and apply feature attribution methods to both visual and quantitative experiments, demonstrating that MetaBreastAI achieves better performance (macro-averaged F1-score: 91.7%; area under the curve: 96.1%) compared to each branch independently, and hybrid baseline models. By highlighting the lesion's location, the heatmap supports interpretability. Due to the explainable detection of lesions in metastatic cases, our hybrid model provides a scalable and clinically feasible strategy through overcoming the downsides of earlier research, then it improves the reliability of AI-assisted techniques in medical decision-making.**

***Index Terms*—Breast cancer metastasis, Convolutional neural network-transformer hybrid, Dual-stream deep learning, Medical image classification, Multi-instance learning.**

ARO-The Scientific Journal of Koya University
Vol. XIV, No.1 (2026), Article ID: ARO.12768. 11 pages
DOI: 10.14500/aro.12768
Received: 2 December 2025; Accepted: 27 March 2026
Regular research paper; Published: 04 June 2026
[†]Corresponding author's e-mail: bnar.ghafour@su.edu.krd

I. Introduction

Nowadays, the presence of a metastatic phenotype significantly complicates clinical decision-making; it leads the breast cancer remains one of the top causes of cancer-related mortality in women globally. The manual interpretation of scans by radiological staff is time-consuming; therefore, automation in diagnosis is being developed. Whereas traditional convolutional neural networks (CNNs) have made significant strides in breast cancer lesion classification, they struggle to model global dependencies (Abdollahi, et al., 2022, Abhisheka, Biswas, and Purkayastha, 2023). Although transformer-based models can learn long-range dependencies (Botlagunta, et al., 2023, Das, et al., 2021), they require large datasets and are generally less locally sensitive than CNNs.

Limited studies have examined hybrid models for breast cancer metastasis detection, weakly labeled datasets, and lack of interpretable frameworks (Allugunti, 2021, Madani, Behzadi and Nabavi, 2022). Furthermore, modalities of the multi-instance learning (MIL) approach are not utilized in existing methods, which are favorable where only coarse-level annotations are available. To address these issues, we propose MetaBreastAI framework that combines a CNN stream enhanced with a convolutional block attention module (CBAM) and a parallel Transformer stream.

Combining an enhanced CNN (CBAM) with a transformer can be used to process medical scans to extract features, which are then fused using a MIL strategy. The main objectives of this study are to develop a novel deep learning architecture that facilitates multi-class metastatic detection, provide clinical transparency through visual explanations, and improve performance to overcome weak supervision issues.

The structure of the paper is as follows: Section 2 provides a review of recent works, Section 3 describes the proposed architecture, Section 4 presents experimental results, performance comparisons, and evaluation. Section 5 discusses findings and limitations. Finally, Section 6 outlines the conclusions and future directions.

II. Related Work

In this section, the prior works on deep learning methods for breast cancer metastasis detection are reviewed.

A. CNN-Based Models for Breast Cancer Detection

Several early systems based on CNNs showed high accuracy in detecting the metastatic patterns from histopathology/whole-slide images (WSIs).

(Abdollahi, et al., 2022) used VGG16 for the detection of lymph node metastases in breast cancer, and achieve an accuracy of 98.84%, confirming the usefulness of fine-tuned pre-trained models. (Das, et al., 2021) combine CNNs with multilayer perceptron that enhances feature generation by applying decomposition methods. (Botlagunta, et al., 2023) developed a machine learning (ML) approach to classify breast cancer metastasis, the model lacks integrated use of image-based modalities and interpretability mechanisms. (Allugunti, 2021) Achieved 99.67% accuracy with CNNs, but applied it to thermographic breast images with even better results than other ML models. (Ahmad, et al., 2022) presented a hybrid model with AlexNet and gated recurrent units for detecting lymph node metastasis, reaching the accuracy of 99.50% on the PCam dataset. (Hu, et al., 2023) highlighted the shift from CNN to transformer-based neural networks in the context of WSI analysis. While recent methods have made advancements, they still struggle to capture global spatial dependencies and long-range interactions. This has led to the use of transformer models and the attention mechanisms.

B. Transformer Architectures and Vision Transformers (ViTs)

ViTs have adapted transformer-based architectures to the computer vision domain (Shakarami, et al., 2023). Proposed a transformer CNN for WSIs classification. The model demonstrated exemplary performance in adapting to changing lighting and staining conditions by incorporating spatial transformer layers into EfficientNet blocks. (Hossain, et al., 2023) despite observer bias in the manual annotation, ViTs were used for automatic Region of Interest detection in HER2 grading with an accuracy of 99%. (Zheng, et al., 2022) using contrastive learning and graph construction, proposed a Graph-Transformer Pipeline to classify subtypes of lung cancer from WSIs, achieving 91% accuracy. (Springenberg, et al., 2023) showed that ViTs outperform modern CNNs across five datasets. (Fu, et al., 2022) suggested a hybrid CNN-Transformer model that focuses on gastric pathology images and yields high classification scores across various image types in multiple datasets. Similarly, (Ikromjanov, et al., 2022) Using attention mechanisms, have shown that ViTs for prostate cancer Gleason grading can help better focus on diagnostically relevant patterns in WSI. (Shao, et al., 2021) shows that transformer-based MIL can take advantage of both morphological and spatial dependencies for WSI classification. They achieved results up to 98.82% area under the curve (AUC). (Ramirez-Mena, et al., 2023) In gene expression of prostate cancer, we used explainable artificial intelligence (AI) for prediction. . (Wani, Kumar, and Bedi, 2024) Introduced XAI methods in the form of DeepXplainer as an interpretable deep learning framework for lung cancer detection. This is a promising approach for medical imaging, but has not yet been investigated in breast histopathological applications that require localized interpretation (Sun, et al., 2023).

C. MIL Techniques

MIL is particularly suitable for large histopathology datasets as it allows models to learn from "bags" – of instances without requiring instance-level labels (Afonso, et al., 2024).

(Shao, et al., 2021) proposed a correlated MIL framework based on a transformer that can capture inter-instance dependencies on different datasets. (Zheng, Jiang, and Yao 2024) utilized a contrastive feature aggregation-based reinforcement learning to select informative instances for generalization. (Sun, et al., 2023, Zheng, Jiang and Yao, 2024, Jin, et al., 2025) aligns the instance- and bag-level labels via a curriculum learning strategy and supervised contrastive loss. Other innovations include multimodal classification with pathology-attention based MIL frameworks (Fu, et al., 2025), attention-based multi-scale MIL (Wibawa, 2022) weakly supervised MIL system for cancer mutation detection (Teramoto, et al., 2021), mirror a movement towards context-sensitive, and explainable MIL systems suited to high-resolution data.

D. Explainable AI in Breast Cancer and Medical Imaging

Despite achieving impressive performance, the poor interpretability of deep learning models poses challenges to the clinical utilization. Due to the importance of transparency and accountability for automated decision-making, Explainable AI (XAI) methods have increasingly been combined into breast cancer diagnostics (Abdollahi, et al., 2022, Abhisheka, Biswas, and Purkayastha, 2023). XAI integration into healthcare applications has received increasing attention from researchers (Keyl, et al., 2022; Loh, et al., 2022).

(Chakraborty, et al., 2021) employed XAI to interpret the key conditions associated with a good prognosis of breast cancer patients. (Jiang and Xu, 2022) proposed a framework with attention-based visualization to localize discriminative areas for lung cancer, making transferable methods to breast cancer critical. (Qu, et al., 2022) used XAI to interpret prediction models for breast cancer metastasis.

MetaBreastAI incorporates explainability into CNN and transformer streams, using attention and region attribution to provide grounded evidence for predictions surrounding metastasis.

E. Hybrid and Ensemble Deep Learning Approaches for Metastasis Detection

As neural systems often fail to capture complex interactions among their components, hybrid and ensemble learning are increasingly being adopted.

(Allugunti, 2021) merged an ensemble of CNNs with a multilayer perceptron. It is outperforming single-network

baselines. (Ahmad, et al., 2022) combined AlexNet and gated recurrent units to capture pure spatial and sequential patterns in lymph node classification, with 99.5% accuracy for the PCam Dataset. (Fu, et al., 2022) adopted a CNN-Transformer for gastric pathology, which obtained over 94% accuracy in multiclass tissue classification. A transformer-Graph Attention Network fusion model was proposed by (Sun, et al., 2023) for breast cancer classification from histopathology images. (Jin, et al., 2025) suggested hierarchical instance-bag label alignment exploiting supervision and contrastive learning to boost recognition tasks. A systematic review by (Barbosa, et al., 2024) showed hybrid attention mechanisms and ensemble aggregation strategies always produce the best performance in classification tasks.

MetaBreastAI combines a MIL schema for explainable and accurate breast cancer metastasis detection.

III. Proposed Framework

The proposed model, MetaBreastAI, addresses two key challenges in cancer metastasis detection: (i) the extraction of both local morphological features and global contextual dependencies, and (ii) providing model explainability for enhanced clinical interpretability. MetaBreastAI integrates the contextual modeling of transformers with the spatial learning capabilities of CNNs, then combined with attention mechanisms and a MIL approach, thus the model can effectively identify complicated tumor patterns that are challenging for single-stream networks. The CNN backbone trained and finetuned on breast cancer metastasis histopathology images of CAMELYON16 (Bandi, et al., 2018) – as it contains a larger number of samples that help improve feature learning – then adapted to the Chest computed tomography (CT)-Scan images Dataset from Kaggle (Hany, 2021) by fine tuning to handle domain shift between datasets. The preprocessing procedures are applied to CT scan images, then fed to our model.

To capture localized patterns and highlight important tissues, the first stream combines a CBAM with ResNet-50. The second stream uses a ViT for capturing global contextual information. It splits the image into patches and uses self-attention processes to learn global contextual relationships. An attention-based MIL pooling layer is used to further process the unified embedding created by merging the feature representations from both streams. The final metastasis prediction across target classes is obtained by passing the resultant feature vector through a fully connected classifier head. The framework includes explainability modules for both streams in order to encourage clinical reliability Fig. 1 shows the entire MetaBreastAI workflow.

A. Data Preprocessing

Images were resized, normalized, and augmented to enhance model generalization. Each image was then divided into uniform patches to capture localized features relevant to metastatic patterns.

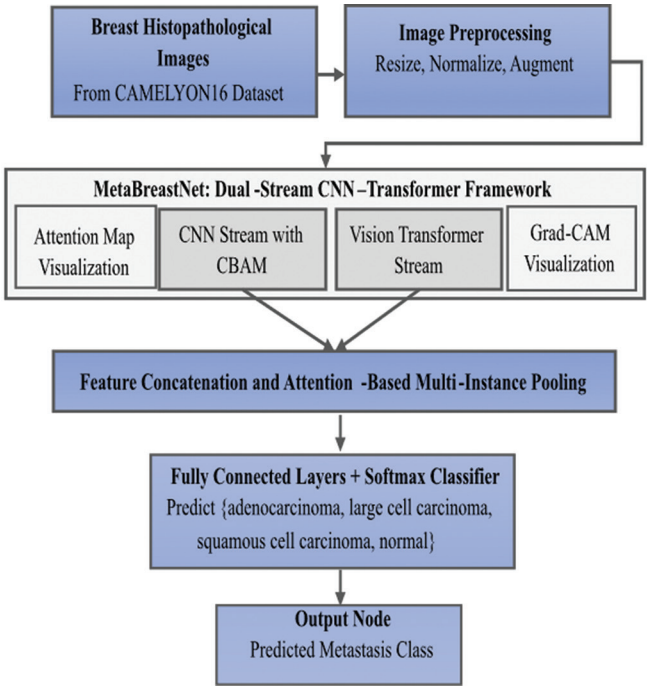

Fig. 1. MetaBreastAI architecture.

Image preprocessing and augmentation

The images were transformed to the RGB color space to be compatible with the pre-trained models and were resized to 224 × 224 × 3 pixels. Data augmentation as random horizontal flipping with a probability of 50%, and random rotation with the angle $\theta \in [-10°, +10°]$, resizing, and normalization were added. The validation and test sets were left untouched to ensure an unbiased evaluation of the model's performance.

Dataset splitting

Due to the relatively small number of images per class in the CT scan dataset, the split ratios were empirically determined. Images were divided into 45% for training, 35% for validation, and 20% for testing.

We used the following equation based on the total number of images N_{total} and subset ratio R as in (1). Where R is the fraction of the dataset given to the respective subset Images

$$N_{subset} = R \times N_{total} \quad (1)$$

B. Meta-BreastNet: Dual-Stream CNN-Transformer

Meta-BreastNet employs a dual-stream architecture, both branches process image patches in parallel, and their outputs are fused to form a unified representation.

CNN stream with CBAM attention

The last average pooling and fully connected layers of the backbone are removed to keep the convolutional layers. A CBAM is integrated with ResNet50, by leveraging both channel-wise and spatial attention mechanisms that incorporated sequentially. CBAM adjusts the feature map and

applies an adaptive average pooling to it, ultimately resulting in a fixed vector by reducing the spatial dimension. The CNN feature vector V_{cnn} reshaped to a signal vector to merge features with the Transformer arm in the latter steps Fig. 2. THESE extracted features enable the CNN stream to broadly capture the important textures and patterns.

ViT stream

The second branch of the MetaBreastNet utilizes a ViT to leverage global contextual relationships across the entire chest image; also, it enables the model to learn spatial relationships between distant patches using self-attention mechanisms. The Transformer stream takes the same resized image, processes it, and converts it into a sequence of patches.

Feature fusion and multi-instance pooling

After obtaining the independent feature vectors from both streams, the next step is the fusion of these complementary features. Finally, the obtained vector is fed to the classification head to output the final prediction.

Classification head

The last layer of MetaBreastNet computes a class prediction by a fully connected classifier followed by ReLU activation, which does the final mapping to the target categories (metastasis). The classification head contains one or more dense layers and dropout regularization to avoid overfitting. The last step is another dense layer with the softmax activation to get the output normalised over the four target classes.

Loss function and optimization

We implemented the cross-entropy loss function for optimization. Where C denotes the number of classes, y_c is ground truth and $\hat{y}_c$ is the predicted probability for class C.

$$L_{CE} = -\sum_{c=1}^{C} y_c \log(\widehat{y_c}) \tag{2}$$

and AdamW optimizer was chosen for parameter optimization in the training process. Where θ is model parameters at iteration t, η is the learning rate, $\widehat{m_t}$ is the bias at first estimation, $\widehat{u_t}$ is the bias at second estimation, λ is the weight decay coefficient and ϵ is a small constant.

$$\theta_{t+1} = \theta_t - \eta . \frac{\widehat{m_t}}{\sqrt{\widehat{u_t} + \epsilon}} - \lambda . \theta_t \tag{3}$$

Training parameters and regularization

Mini-batch size of eight images was employed per iteration. We trained for a maximum of 20 epochs, with 0.0001 learning rate and a dropout layer (0.3) in the classification head, DataLoader in Pytorch was used to load and augment the mini-batch data on the

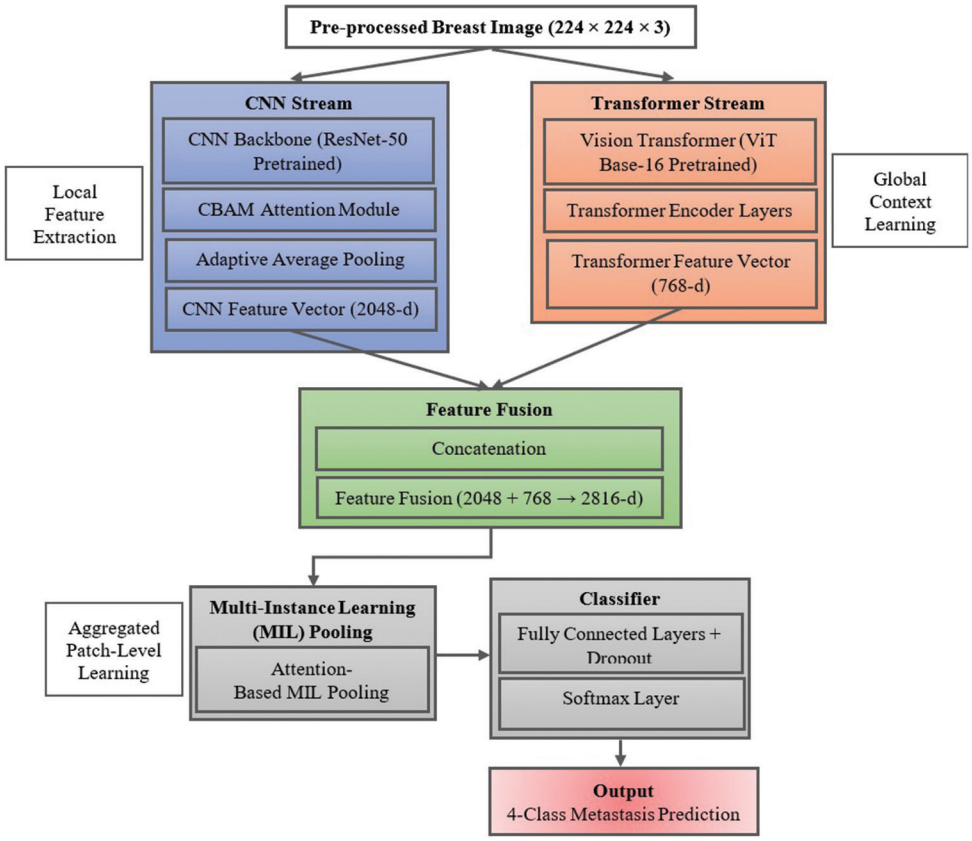

Fig. 2. MetaBreastNet architecture.

fly. The total training time for the CT scan dataset was approximately 4 h.

C. Model Explainability

Gradient-weighted class activation mapping (Grad-CAM) and attention heatmaps are used to enhances interpretability, and verify whether the model focuses on medically relevant metastatic features during prediction.

Grad-CAM for CNN stream

MetaBreastNet generates heatmaps of learned local features by applying Grad-CAM on the CNN stream. It works by calculating the gradient of the output class score for a target class with respect to the feature maps obtained from the CNN stream. These gradients indicate the contribution of each spatial location in the feature map to the final prediction.

Attention map visualization for transformer stream

The MetaBreastNet features Grad-CAM-based explainability for the CNN stream, along with interpretability of the global feature extraction through the inherent self-attention mechanism of the ViT. The attention maps from the Transformer stream represent the pairwise contextual relations across different regions of the input image, providing an overview of how image regions attend to the other areas when predicting for metastasis.

D. Algorithmic Implementation

The pipeline of the MetaBreastNet, including the training and inference process, is depicted in Algorithm 1. After the training is finished, we store the model. Then, a test image is passed through both streams, we extract V_{cnn} and V_{trans}, fuse them as V_{fusion}, and pool them same way as during training. The model predicts the class label $\hat{y}$ for metastasis. Explainability modules can also be applied to explain the predictions the model makes.

Algorithm 1: MetaBreastNet Training And Inference Workflow

Algorithm: MetaBreastNet Training and Inference Workflow
Input: Preprocessed histopathological images I, metastasis class labels y
Output: Predicted class labels $\hat{y}$

1. Initialize MetaBreastNet with pretrained ResNet-50 and Vision Transformer backbones.
2. For each training epoch:
   a. For each mini-batch (I, y)
      i. Apply data augmentation to I.
      ii. Extract CNN features V_{cnn} from I using ResNet-50+CBAM
      iii. Extract Transformer features V_{trans} from I using Vision Transformer.
      iv. Concatenate V_{cnn} and V_{trans} to obtain V_{fusion}.
      v. Apply attention-based MIL pooling to V_{fusion} to obtain V_{mil}.
      vi. Pass V_{mil} through fully connected layers and softmax to obtain $\hat{y}$.
      vii. Compute loss L_{CE}between $\hat{y}$ and y.
      viii. Update model parameters using AdamW optimizer.
3. Save the trained MetaBreastNet model.
4. During inference:
   a. Input test image I.
   b. Extract V_{cnn} and V_{trans} using trained feature extractors.
   c. Fuse features, apply MIL pooling, and predict $\hat{y}$.
   d. Optionally, generate Grad-CAM and attention maps for explainability.
   e. Output predicted class label $\hat{y}$.

IV. Experimental Result

This section describes the setup of the challenge, evaluation metrics, and comparisons with baseline models.

A. Experimental Setup

All experiments were conducted on a laptop equipped with an NVIDIA GeoForce Experience 3.28.0.417, 13th Gen Intel(R) Core (TM) i7-13700HX, 2.10 GHz, running Windows 10. We used a model trained in Python 3.9 with PyTorch 1.13.1, and other supporting libraries were OpenSlide for handling WSI, Albumentations for data augmentation, and Scikit-learn for metrics calculation.

We trained the model for 100 epochs, but with early stopping based on the validation AUC. All experiments were repeated 5 times, with the average performance being reported for statistical significance.

B. Exploratory Data Analysis and Preprocessing

The evaluation of MetaBreastAI was conducted using the CAMELYON16 dataset (Bandi, et al., 2018). Each slide was divided into non-overlapping patches of size 256 × 256 pixels at ×20 magnification. Otsu thresholding and morphological filtering were used to identify tissue regions; non-tissue patches were excluded, and about 250,000 pertinent patches were produced. The input patches to the model were scaled to 224 × 224 pixels. To achieve objective evaluation, the dataset was divided into 70% training (189 WSIs), 10% validation (27 WSIs), and 20% testing (54 WSIs). Each patch normalized by Reinhard stain normalization.In this paper, we adapted our model to CT scan dataset in Fig. 3 based on four classes, which are adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal tissues, as shown in Fig. 4. These examples emphasize the inter-class visual heterogeneity and structural differences.

Augmented CT samples denoting sample CT scan for different classes are shown in Fig. 5.

C. Performance Metrics

Accuracy, Precision, Recall, and F1-score, AUC, the receiver operating characteristic curve (AUC-ROC) were calculated. We computed a multi-class confusion matrix (Fig. 6) to facilitate comparison of predictions with the ground truth.

D. Proposed Model Results

The average classification performance across all four classes exceeded 91% Table I. ROC, and precision-recall curves confirmed the same results.

Attention heatmaps confirmed that the model focused

http://dx.doi.org/10.14500/aro.12768

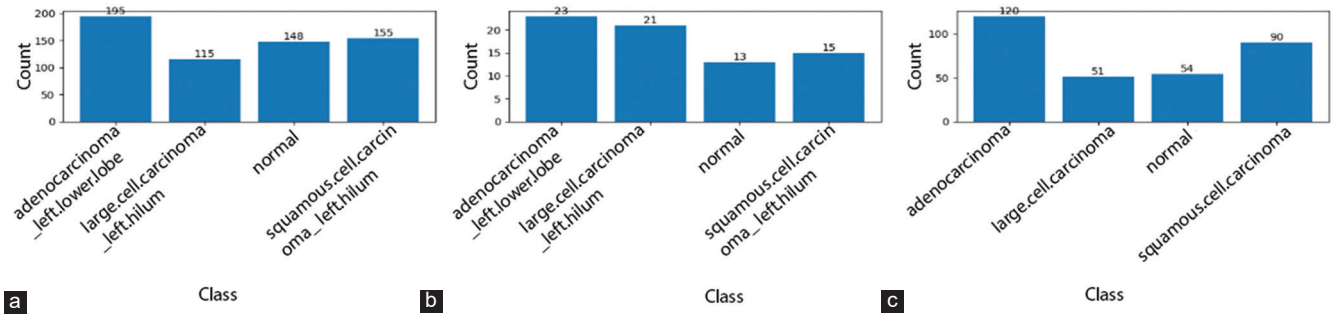

Fig. 3. Dataset class distributions across train, validation, and test splits. (a) Train set class distribution (b) Validation set class distribution (c) Test set class distribution.

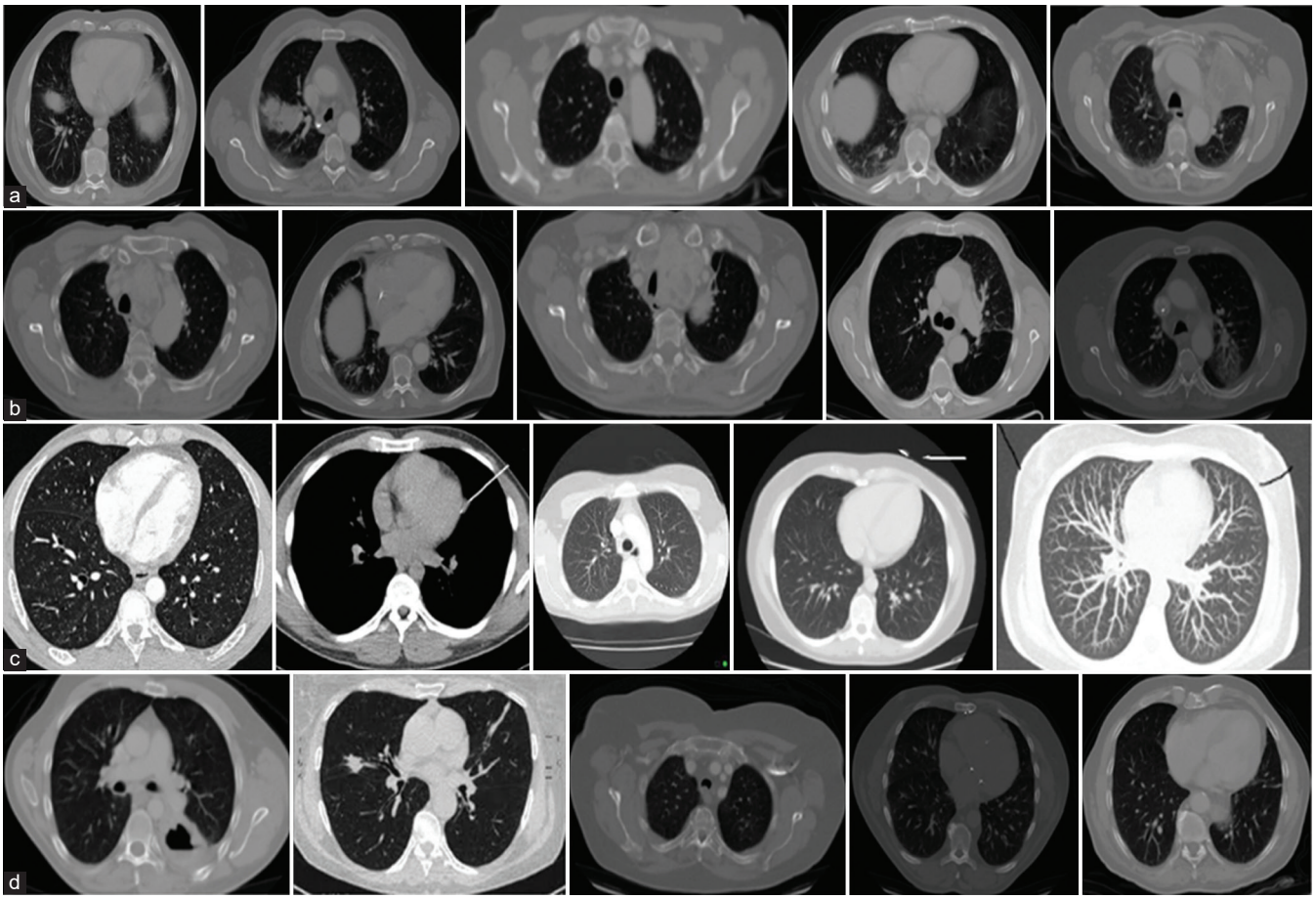

Fig. 4. Representative computed tomography scan samples from each class (a) Adenocarcinoma_left.lower.lobe (b) Large cell_carcinoma_left.hilum (c) Normal (d) squamous_cell_carcinoma_left.hilum_T1_N2_M0_IIIa.

TABLE I
PER-CLASS CLASSIFICATION PERFORMANCE OF METABREASTAI ON THE TEST SET

Class Label	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Adenocarcinoma	94.6	92.3	93.4	96.2
Large Cell Carcinoma	90.2	87.1	88.6	94.1
Squamous Cell Carcinoma	92.8	94.1	93.4	96.7
Normal Tissue	95.1	93.9	94.5	97.4
Macro Average	93.2	91.85	91.7	96.1

on regions that were relevant to the lesions. The results demonstrate the framework's robustness, reliability, and clinical applicability in detecting mass location from CT scans after presenting the result to a radiologist. The model's precision for the Normal Tissue was the greatest at 95.1%, indicating a low false positive rate. Squamous_cell_carcinoma had the highest recall (94.1%), demonstrating exceptional sensitivity in identifying this subtype. With the Normal class achieving the highest F1-Score of 94.5%,

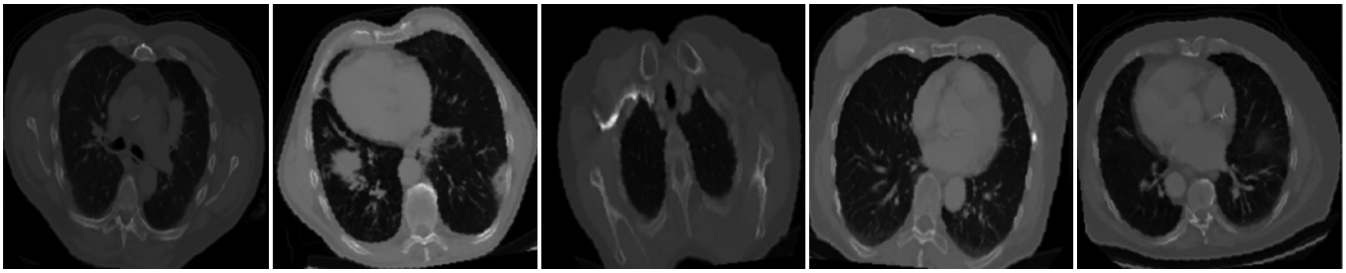

Fig. 5. Augmented computed tomography scan samples representing each class.

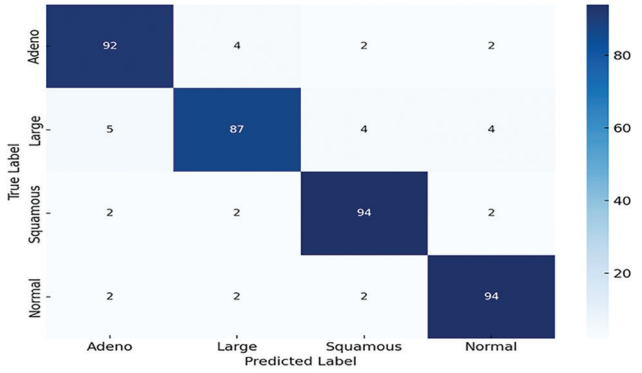

Fig. 6. Confusion matrix of MetaBreastAI on multiclass classification.

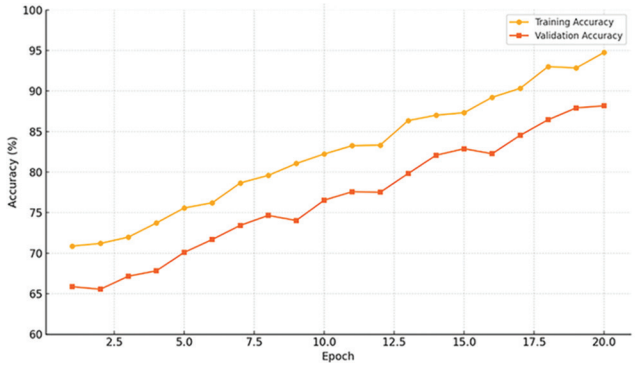

Fig. 7. Training and validation accuracy dynamics of MetaBreastAI.

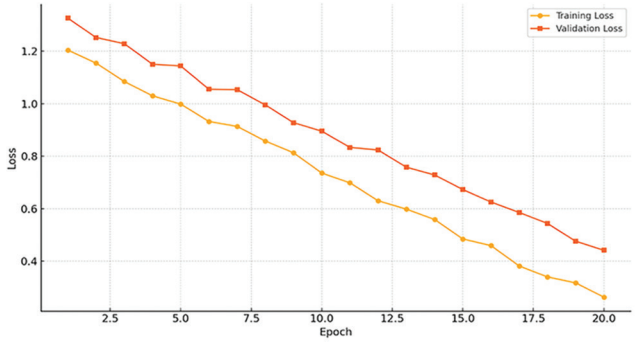

Fig. 8. Training and validation loss dynamics of MetaBreastAI.

and Adenocarcinoma and Squamous_Cell_Carcinoma both achieving 93.4%. Large_Cell_Carcinoma's F1-score was lower at 88.6%, which is consistent with earlier results of decreased sensitivity and highlights the need for better discriminating in this group. The training and validation accuracy of MetaBreastAI are displayed in Fig. 7. The model closely follows the training curve and shows steady convergence, with validation accuracy reaching about 96%. This stability is consistent with the great classification accuracy and shows how effectively the model generalizes. The training and validation loss curves over 20 epochs are displayed in Fig. 8. Stable convergence and useful model training are indicated by the losses gradually decreasing with little divergence between the two curves.

Fig. 6 demonstrates MetaBreastAI's confusion matrix. Minor misclassifications are between large cell carcinoma and adenocarcinoma.

E. Baseline Comparison and the Ablation Study

A number of baseline models were assessed using the same experimental conditions in order to verify the efficiency of the MetaBreastAI. VGG16, ResNet50, and EfficientNet-B0 were included. These models showed limited capacity to effectively capture the local and global features necessary for differentiating between closely related cancer subtypes. With an F1-score of 88.1% and an AUC of 93.5%, EfficientNet-B0 outperformed VGG16 and ResNet50, but it is still behind the suggested model (Table II). To evaluate their impact on classification performance, two ablation baselines were also taken into consideration, number of components was eliminated. While the Transformer branch used a Swin Transformer module without CNN, the CNN-only branch used a ResNet-based backbone without the Transformer. With an F1-score of 89.6% for the CNN-only stream and 89.0% for the Transformer-only stream, both demonstrated better performance than classical CNNs, especially in recall and AUC metrics. However, they still underperformed compared to the complete MetaBreastAI. With 93.2% precision, 91.85% recall, 91.7% F1-score, and 96.1% AUC, MetaBreastAI outperformed all in Table III. This confirms the value of our model in addressing the complex nature of metastatic breast cancer image classification. Every metric is outperformed by MetaBreastAI.

These findings emphasize how crucial it is to combine convolutional and self-attention-based architectures in order to effectively classify complicated metastatic cancer subtypes.

F. Explainability and Visual Interpretation

Grad-CAM by matching model attention with diagnostically significant locations in the CT scans successfully localized metastatic regions across various cancer subtypes. The attention heatmap in Fig. 9 shows concentrated activation

at the lesion area that has been annotated. The model's dependability and diagnostic transparency are strengthened by this interpretability, as it was confirmed by a radiologist with more than 17 years of experience.A misclassification of MetaBreastAI, for a CT scan categorized as adenocarcinoma as large cell carcinoma, is shown in (Fig. 9b). The Grad-CAM layer displays strong activation in lung zones, whereas the attention heatmap displays scattered focus regions. This highlights the model uncertainty in identifying subtle characteristics in metastatic progression. (Fig. 9c) illustrates a successful classification of squamous cell carcinoma. The tumor's location shows strong activation in the attention heatmap. A true prediction for a typical tissue CT scan is shown in (Fig. 9d). The Grad-CAM indicates scattered, low-intensity focus throughout non-suspicious regions, whereas the attention heatmap shows minor activation. It successfully prevents mistaken localization in the absence of diseased inputs. (Fig. 9e) shows an example of adenocarcinoma that accurately identified. Both the Grad-CAM overlay and the attention heatmap show strong activation in the location of the tumor. This illustrates its interpretability in identifying critical regions suggestive of metastatic adenocarcinoma in the diagnosis of breast cancer.

TABLE I
Per-class Classification Performance of MetaBreastAI on the Test Set

Class Label	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Adenocarcinoma	94.6	92.3	93.4	96.2
Large Cell Carcinoma	90.2	87.1	88.6	94.1
Squamous Cell Carcinoma	92.8	94.1	93.4	96.7
Normal Tissue	95.1	93.9	94.5	97.4
Macro Average	93.2	91.85	91.7	96.1

TABLE III
Performance Impact of Component Removal in Ablation Study

Configuration	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Full MetaBreastAI	93.2	91.85	91.7	96.1
Without Transformer Stream	90.4	88.1	88.6	93.7
Without CNN Stream	89.5	87.0	87.9	93.1
WithoutMulti-Instance Pooling	91.1	89.3	89.6	94.2
Without the Attention Module	91.5	89.6	90.2	94.6

G. External Validation and Generalization Analysis

We performed external validation using the public TCGA-BRCA dataset (Tomczak, Czerwińska, and Wiznerowicz, 2015) to assess MetaBreastAI's robustness and generalizability. It showed consistent and dependable in spite of data variation. In order to simulate real-world deployment, the external evaluation evaluates the model's capacity to accurately categorize metastatic subtypes using the pre-trained model without fine-tuning. In addition, we conducted a domain adaptation experiment in which batch normalization updates were utilized to calibrate a small fraction (10%) of the TCGA-BRCA dataset. The model's capacity to adjust to the domain followed modest gains in precision and recall.

Table IV provides a comparative summary of performance metrics across internal and external datasets.With an F1-score of 94.5% internally and 93.1% externally, the model performs great in the Normal Tissue class on both datasets. A slight decline from 93.4% to 91.2% is noted for the adenocarcinoma class, indicating a modest vulnerability to domain differences. With only a ~2.7% drop (93.4% internal vs. 90.7% external), the squamous cell carcinoma class has strong consistency. Large cell carcinoma had the biggest F1-score decline, from 88.6% to 86.3%, suggesting that this subtype is more susceptible to inter-dataset variability. A similar pattern may be seen with the AUC metric. Normal Tissue achieves the highest AUC (97.4% internal vs. 96.4% external), and squamous cell carcinoma maintains excellent AUC values (96.7% internal vs. 95.3% exterior). AUC values for Adenocarcinoma and Large Cell Carcinoma show minor decreases (from 96.2% to 94.7% and from 94.1% to 92.8%, respectively), but they are still within clinically acceptable ranges.

The graphical comparison reveals that MetaBreastAI only slightly drops in external validation while maintaining good classification performance across datasets presented in Fig. 10.

H. Statistical Significance and Confidence Interval Analysis

We performed a statistical analysis for performance comparisons and the estimation of confidence intervals (CIs) for evaluation metrics to guarantee the stability of MetaBreastAI across various test samples.

We used the bootstrapping method with 1,000 resamples and calculated 95% CIs for F1-Score and AUC. The mean MetaBreastAI improvements over all five baseline models are

TABLE IV
Cross-Dataset Performance Comparison of MetaBreastAI

Class Label	Dataset	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Adenocarcinoma	Internal	94.6	92.3	93.4	96.2
	External (TCGA)	91.2	88.4	89.7	94.5
Large Cell Carcinoma	Internal	90.2	87.1	88.6	94.1
	External (TCGA)	88.1	84.9	86.4	92.6
Squamous Cell Carcinoma	Internal	92.8	94.1	93.4	96.7
	External (TCGA)	90.5	91.3	90.9	95.1
Normal Tissue	Internal	95.1	93.9	94.5	97.4
	External (TCGA)	92.6	90.7	91.6	95.8
Macro Average	Internal	93.2	91.85	91.7	96.1
	External (TCGA)	90.6	88.83	89.65	94.5

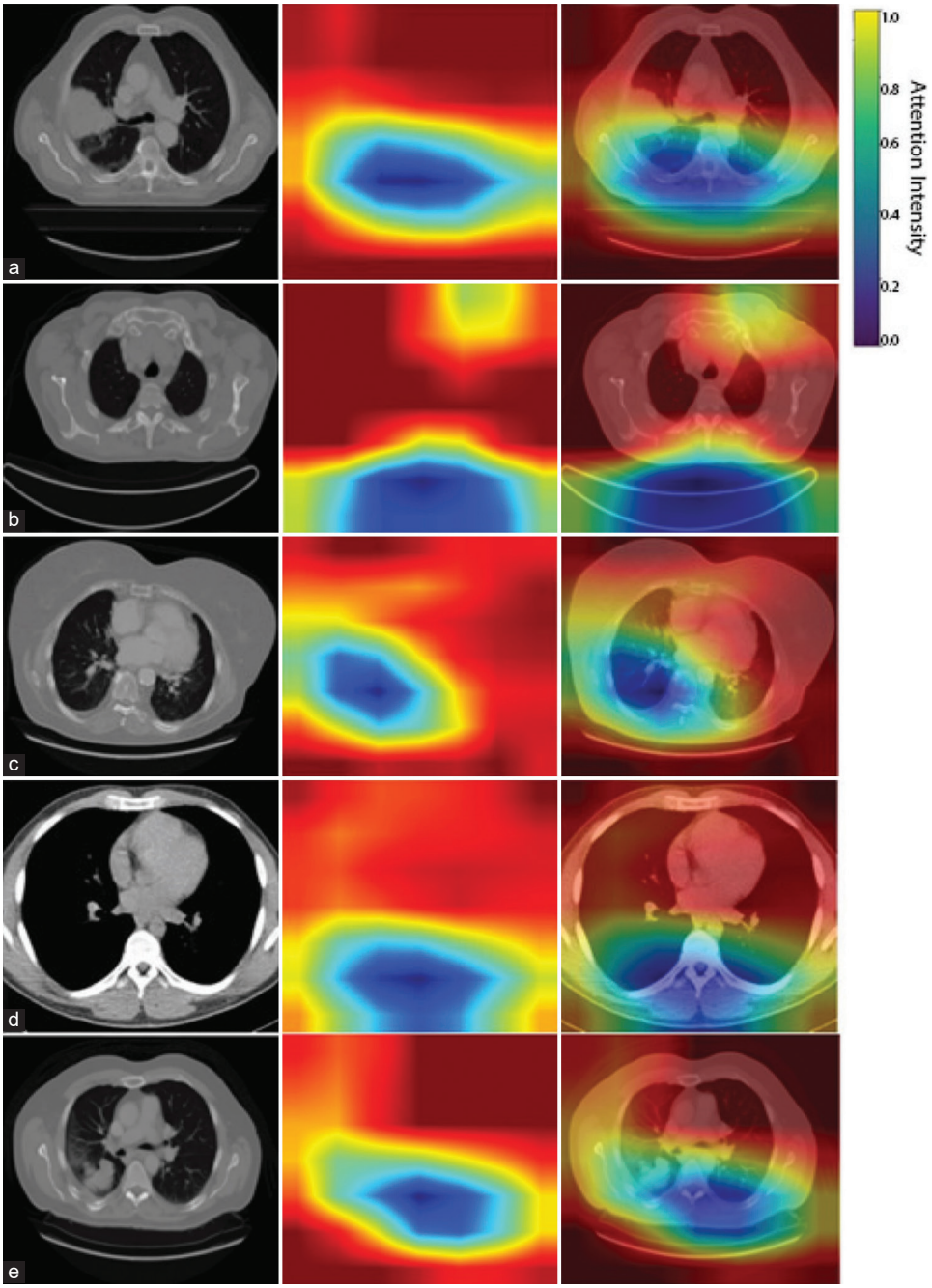

Fig. 9. Grad-CAM visualization for (a) Adenocarcinoma detection (b) Misclassified computed tomography scan sample (c) Correctly classified squamous cell carcinoma (d) Correctly classified normal tissue sample (e) Correctly classified adenocarcinoma sample.

TABLE V
CONFIDENCE INTERVALS FOR PERFORMANCE METRICS OF METABREASTAI

Metric	Mean (%)	95% Confidence Interval (%)
F1-Score	91.7	[90.1, 93.2]
AUC	96.1	[95.0, 97.3]
Precision	93.2	[91.6, 94.7]
Recall	91.85	[90.3, 93.1]

shown in Table V. We performed the Wilcoxon signed-rank test on 5-fold cross-validation outputs for both metrics to determine whether the improvements are statistically significant or within standard variation. The test gave p = 0.003 for the AUC and 0.006 for the F1-score, which are more than 4 times below the standard significance threshold of 0.05. This confirms that MetaBreastAI's gain in classification ability is statistically significant and not the result of chance.

V. DISCUSSION

Particularly in high-impact diagnostic applications, the identification of breast cancer metastases is a major computational pathology challenge. The earlier CNNs and Transformer frequently lack interpretability and have trouble integrating spatial and global context. We introduce

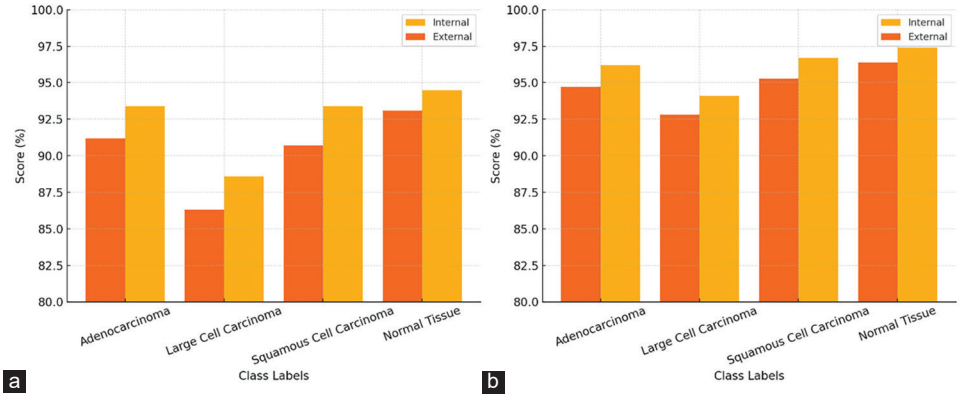

Fig. 10. Comparative evaluation of generalization performance (a) F1-score (b) Area under the curve comparison.

MetaBreastAI, a novel dual-stream deep learning architecture that combines CNN with Transformer and CBAM attention in an MIL context, to overcome these drawbacks. This methodology provides an interpretable approach to multi-class metastatic detection while addressing the limitations of current models, including restricted generalizability, poor interpretability, by dividing the duties of feature improvement and aggregation into two different processes.

There are two limitations of the present study. One, the model was developed and tested primarily on a small number of publicly available CT scan datasets, which may not adequately reflect real-world clinical heterogeneity. Second, the use of attention maps improves interpretability; however, it does not involve clinical validation by expert radiologists, which would provide insights into practical deployment.

VI. Conclusion And Future Work

The proposed system is a dual-stream explainable model that can identify the location of cancer cells in CT scans. It utilizes the combined capabilities of CNNs and Transformer blocks for global context modeling.

The results of this study are encouraging, but it is limited by the lack of clinical validation and the small size of the sample. Further studies should focus on multi-institutional and real-world datasets that can reflect the variability of the environment.

The development of the CNN-Transformer system will be instrumental in establishing its capability as a decision support tool for oncologic imaging.

VII. References

Abdollahi, J., Davari, N., Panahi, Y., and Gardaneh, M., 2022. Detection of metastatic breast cancer from whole-slide pathology images using an ensemble deep-learning method. *Archives of Breast Cancer*, 9, pp.364-376.

Abhisheka, B., Biswas, S.K., and Purkayastha, B., 2023. A comprehensive review on breast cancer detection, classification and segmentation using deep learning. *Archives of Computational Methods in Engineering*, 30, pp.5023-5052.

Afonso, M., Bhawsar, P.M.S., Saha, M., Almeida, J.S., and Oliveira, A.L., 2024. Multiple instance learning for WSI: A comparative analysis of attention-based approaches. *Journal Pathology Informatics*, 15, p.100403.

Ahmad, S., Ullah, T., Ahmad, I., Al-Sharabi, A., Ullah, K., Khan, R.A., Rasheed, S., Ullah, I., Uddin, M.N., and Ali, M.S., 2022. A novel hybrid deep learning model for metastatic cancer detection. *Computational Intelligence and Neuroscience,* 2022, p.8141530.

Allugunti, V.R., 2021. Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. *International Journal of Engineering in Computer Science*, 41, pp.49-56.

Bandi, P., Geessink, O., Manson, Q., Van Dijk, M., Balkenhol, M., Hermsen, M., Bejnordi, B.E., Lee, B., Paeng, K., Zhong, A., Li, Q., Zanjani, F.G., Zinger, S., Fukuta, K.,… & Litjens, G., 2018. From detection of individual metastases to classification of lymph node status at the patient level: The CAMELYON17 challenge. *IEEE Transactions on Medical Imaging*, 38, pp.550-560.

Barbosa, D., Ferreira, M., Junior, G.B., Salgado, M., and Cunha, A., 2024. Multiple instance learning in medical images: A systematic review. *IEEE Access,* 12**,** pp.78409-78422.

Botlagunta, M., Botlagunta, M.D., Myneni, M.B., Lakshmi, D., Nayyar, A., Gullapalli, J.S., and Shah, M.A., 2023. Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms. *Scientific Reports,* 13, p.485.

Chakraborty, D., Ivan, C., Amero, P., Khan, M., Rodriguez-Aguayo, C., Basagaoglu, H., and Lopez-Berestein, G., 2021. Explainable artificial intelligence reveals novel insight into tumor microenvironment conditions linked with better prognosis in patients with breast cancer. *Cancers (Basel),* 13, p.3450.

Das, A., Narayan Mohanty, M., Kumar Mallick, P., Tiwari, P., Muhammad, K., and Zhu, H., 2021. Breast cancer detection using an ensemble deep learning method. *Biomedical Signal Processing and Control,* 70, p.103009.

Fu, B., Zhang, M., He, J., Cao, Y., Guo, Y., and Wang, R., 2022. StoHisNet: A hybrid multi-classification model with CNN and Transformer for gastric pathology images. *Computer Methods and Programs in Biomedicine,* 221**,** p.106924.

Fu, F., Zhang, X., Wang, Z., Xie, L., Fu, M., Peng, J., Wu, J., Wang, Z., Guan, T., He, Y., Lin, J.S., Zhu, L., and Dai, W., 2025. A pathology-attention multi-instance learning framework for multimodal classification of colorectal lesions. *Frontiers Pharmacology,* 16**,** p.1592950.

Hany, M., 2021. *Chest CT-Scan Images Dataset*. Kaggle, United States.

Hossain, M.S., Shahriar, G.M., Syeed, M.M.M., Uddin, M.F., Hasan, M., Shivam, S., and Advani, S., 2023. Region of interest (ROI) selection using vision transformer for automatic analysis using whole slide images. *Scientific Reports,* 13, p.11314.

Hu, W., Li, X., Li, C., Li, R., Jiang, T., Sun, H., Huang, X., Grzegorzek, M., and Li, X., 2023. A state-of-the-art survey of artificial neural networks for whole-slide image analysis: From popular convolutional neural networks to potential visual transformers. *Computers in Biology and Medicine,* 161**,** p.107034.

Ikromjanov, K., Bhattacharjee, S., Hwang, Y.B., Sumon, R.I., Kim, H.C., and

Choi, H.K., 2022. Whole Slide Image Analysis and Detection of Prostate Cancer using Vision Transformers. In: *2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC).*

Jiang, X., and XU, C., 2022. Deep learning and machine learning with grid search to predict later occurrence of breast cancer metastasis using clinical data. *Journal Clinical Medicine,* 11, p.5772.

Jin, C., Luo, L., Lin, H., Hou, J., and Chen, H., 2025. HMIL: Hierarchical multi-instance learning for fine-grained whole slide image classification. *IEEE Transactions on Medical Imaging,* 44, pp.1796-1808.

Keyl, P., Bockmayr, M., Heim, D., Dernbach, G., Montavon, G., Muller, K.R., and Klauschen, F., 2022. Patient-level proteomic network prediction by explainable artificial intelligence. *NPJ Precision Oncology,* 6**,** p.35.

Loh, H.W., Ooi, C.P., Seoni, S., Barua, P.D., Molinari, F., and Acharya, U.R., 2022. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011-2022). *Computer Methods and Programs in Biomedicine,* 226, p.107161.

Madani, M., Behzadi, M.M., and Nabavi, S., 2022. The role of deep learning in advancing breast cancer detection using different imaging modalities: A systematic review. *Cancers (Basel),* 14, p.5334.

Qu, L., Liu, S., Liu, X., Wang, M., and Song, Z., 2022. Towards label-efficient automatic diagnosis and analysis: A comprehensive survey of advanced deep learning-based weakly-supervised, semi-supervised and self-supervised techniques in histopathological image analysis. *Physics Medicine Biology,* 67, p.20TR01.

Ramirez-Mena, A., Andres-Leon, E., Alvarez-Cubero, M.J., Anguita-Ruiz, A., Martinez-Gonzalez, L J., and Alcala-Fdez, J., 2023. Explainable artificial intelligence to predict and identify prostate cancer tissue by gene expression. *Computer Methods and Programs in Biomedicine,* 240**,** p.107719.

Shakarami, A., Nicolè, L., Terreran, M., Paolo Dei Tos, A., and Ghidoni, S., 2023. TCNN: A Transformer Convolutional Neural Network for artifact classification in whole slide images. *Biomedical Signal Processing and Control,* 84, p.104812.

Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., and Zhang, Y., 2021. *TransMIL: Transformer based Correlated Multiple Instance Learning for Whole Slide Image Classification*. Cornell University, New York.

Springenberg, M., Frommholz, A., Wenzel, M., Weicken, E., Ma, J., and Strodthoff, N., 2023. From modern CNNs to vision transformers: Assessing the performance, robustness, and classification strategies of deep learning models in histopathology. *Medical Image Analysis*, 87, p.102809.

Sun, X., Li, W., Fu, B., Peng, Y., He, J., Wang, L., Yang, T., Meng, X., Li, J., Wang, J., Huang, P., and Wang, R., 2023. TGMIL: A hybrid multi-instance learning model based on the transformer and the graph attention network for whole-slide images classification of renal cell carcinoma. *Computer Methods Programs Biomedicine***,** 242, p.107789.

Teramoto, A., Kiriyama, Y., Tsukamoto, T., Sakurai, E., Michiba, A., Imaizumi, K., Saito, K., and Fujita, H., 2021. Weakly supervised learning for classification of lung cytological images using attention-based multiple instance learning. *Scientific Reports,* 11, p.20317.

Tomczak, K., Czerwińska, P., and Wiznerowicz, M., 2015. The cancer genome atlas (TCGA): An immeasurable source of knowledge. *Contemporary Oncology* (*Pozn*), 19, pp.A68-A77.

Wani, N.A., Kumar, R., and Bedi, J., 2024. DeepXplainer: An interpretable deep learning based approach for lung cancer detection using explainable artificial intelligence. *Computer Methods and Programs in Biomedicine,* 243, p.107879.

Wibawa, M.S., Lo, K.W., Young, L.S., and Rajpoot, N., 2022. *Multi-Scale Attention-based Multiple Instance Learning for Classification of Multi-Gigapixel Histology Images*. Cornell University, New York, pp.635-647.

Zheng, T., Jiang, K., and Yao, H., 2024. *Dynamic Policy-Driven Adaptive Multi-Instance Learning for Whole Slide Image Classification*. IEEE, Vancouver, pp.8028-8037.

Zheng, Y., Gindra, R.H., Green, E.J., Burks, E.J., Betke, M., Beane, J.E., and Kolachalama, V.B., 2022. A graph-transformer for whole slide image classification. *IEEE Transactions on Medical Imaging,* 41, pp.3003-3015.