IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 February 2026, accepted 3 May 2026, date of publication 25 May 2026, date of current version 28 May 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3691800

RESEARCH ARTICLE

Deep Learning Framework for Breast Tumor Detection and Segmentation Using CT Imaging

ZAINAB PERVAIZ[1], TALLAT JABEEN[2], RUQAIYYA ADIL[3], MUHAMMAD WALEED KHAN[1], AAYSHA TARIQ[4], MUHAMMAD IRFAN ZAFAR[5], AND ISHRAT JABEEN[1]

[1]School of Interdisciplinary Engineering and Sciences (SINES), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan
[2]Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia
[3]NUST School of Health Sciences, Islamabad 44000, Pakistan
[4]Pakistan Atomic Energy Commission Hospital (PAEC), Islamabad 44000, Pakistan
[5]School of Mechanical and Manufacturing Engineering (SMME), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan

Corresponding author: Ishrat Jabeen (ishrat.jabeen@sines.nust.edu.pk)

This work was conducted without any research fundings. Research facilities of the Computational Drug Design Lab (CDD) at NUST, Islamabad, Pakistan was used to conduct this study.

ABSTRACT Breast cancer remains the leading malignancy among women globally, yet AI-based detection systems focus predominantly on mammography while whole-body CT imaging is routinely performed for oncological staging thus, remains critically underexplored. Existing CT-based studies are limited by small datasets, geographical bias, and lack of spatial localization capabilities. This study introduces a radiologist-verified CT scan dataset of 9330 annotated slices from three Pakistani hospitals, approved by institutional ethics committees. Data was partitioned into training (70%), validation (15%), and test (15%) sets to develop a two-stage deep learning framework. The training data includes 6,106 whole-body CT slices (3,053 tumor, 3,053 normal). The perfectly balanced training dataset eliminates class-level imbalance, while detection-level imbalance is addressed through distributional focal loss mechanisms. We systematically evaluated seven transfer learning architectures (VGG16, VGG19, ResNet50V2, InceptionV3, DenseNet121, DenseNet201, EfficientNetB0–B5), revealing that VGG16 achieved optimal performance (82% accuracy, 0.80 AUC-ROC) compared to deeper architectures (ResNet50V2: 49%, EfficientNet: 50%), demonstrating that simpler models exhibit better inductive bias for constrained medical imaging datasets. Building upon this, we developed an integrated YOLOv11-Seg framework exploiting architectural innovations: C3K2 bottleneck modules for hierarchical feature extraction, Spatial Pyramid Pooling Fast (SPPF) for multi-scale context aggregation, and Pixel Spatial Attention (C2PSA) for enhanced boundary delineation. YOLOv11-Seg achieved 84.6% precision, 90% recall, 0.84 mAP@50, 0.839 Dice coefficient, and 0.724 mean IoU. Error analysis revealed false negatives (3.83%) primarily from low-contrast lesions and small tumors ($<$8 mm), while false positives (6.3%) resulted from fibrous tissue misclassification. Real-time video inference on $>$10,000 frames demonstrated 10 FPS processing, 93.1% sensitivity, 97.9% specificity, and 94.3% temporal consistency, enabling clinical workflow integration. Limitations of the current study include single-radiologist annotation, geographical homogeneity, retrospective design, and binary classification scope, warranting prospective multi-center validation, multi-class classification extension, and explainable AI development. This work establishes foundation for AI-assisted opportunistic breast cancer screening in whole-body CT imaging, particularly valuable for resource-constrained settings, with clear pathways toward comprehensive staging and diagnostic support systems.

INDEX TERMS Breast cancer detection, CT imaging, instance segmentation, deep learning, computer aided-detection, real-time tumor detection, YOLOv11-Seg.

I. INTRODUCTION

In recent years, breast cancer (BC) remains the leading malignancy among women globally, with 2.3M new cases

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Ye.

© 2026 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

and approximately 700,000 deaths annually [1], [2]. While mammography have been the frontline [3] and gold standard for screening [4], but mammography-based screening has reduced mortality rates over three decades [5], persistent limitations in sensitivity (30-48% in dense breast tissue) and specificity (7-12% false-positive (HU windowing) rates) continue to challenge early detection. false reassurance [6], [7]. These limitations are amplified in low- and middle-income countries (LMICs) where resource constraints [8] and radiologist shortages further impede cancer control [9]. The increasing imaging workload with radiologists [10] interpreting 50-100 cases daily [11], exacerbates diagnostic errors [12] due to fatigue and inter-observer variability [13], [14], creating urgent demand for automated diagnostic support system [15], [16].

Computed tomography (CT) has emerged as a promising three-dimensional imaging modality that addresses mammography's inherent limitations [17]. Unlike projection-based mammography, CT acquires isotropic volumetric data, eliminating tissue superposition and enabling superior visualization of lesion morphology, margins, and spatial relationships [18], [19], [20]. Studies demonstrate improved contrast resolution, enhanced detection of architectural distortions, and more accurate tumor size assessment compared to conventional imaging [21], [22], [23].

In current clinical practice, the diagnostic pathway typically involves initial screening with mammography or ultrasound, followed by whole-body CT scans for patients with confirmed aggressive tumors to assess metastatic spread, lymph node involvement, and distant organ invasion [24]. However, these CT scans already acquired for staging purposes contain rich information about the primary tumor that remains underutilized for diagnostic purposes. This presents a unique opportunity: leveraging the same CT imaging performed for staging to simultaneously detect and characterize breast lesions, thereby maximizing the diagnostic yield from a single imaging examination. Such an integrated approach could be particularly valuable in resource-constrained settings where multiple imaging modalities may not be readily available, and in cases where patients present with advanced disease requiring both detection and staging assessment. Despite these potential advantages, clinical adoption of CT-based BC detection remains limited due to radiation concerns [25], equipment costs, and critically, the absence of robust computer-aided detection (CAD) systems specifically designed for breast CT analysis [26]. This gap represents a significant barrier to realizing CT's diagnostic potential and an opportunity for advanced artificial intelligence (AI) methods.

Deep Learning (DL) have revolutionized medical image analysis, with convolutional neural network (CNN) achieving human-level performance in lesion detection and classification [26]. Despite decades of success in mammography [27], [28],) [29], [30], [31], [32], [33], AI applications for CT-based BC detection remain hindered by scarce annotated data [34], [35], [36], [37], [38], data standardization [39], limited generalization, and integration challenges [40]. Recent comprehensive reviews highlight substantial progress in breast imaging AI, with DL models demonstrating 5-15% improvements in cancer detection rates and up to 20% reductions in false negatives (FNs) across mammography, ultrasound, MRI [36]. State-of-art approaches now explainable AI frameworks and foundation models to enhance clinical interpretability and generalization [12], [41].

II. RELATED WORKS

Most of the previous research focuses on mammography [42], [43], [44], [45], with recent YOLO-based detection frameworks achieving impressive real-time performance [46], [47], [48]. These studies frequently suffer from limited dataset sizes ($<$1000 cases), lack of demographic diversity, and poor external validation. Similar constraints affect ultrasound and MRI studies [49] [50], [51], where overfitting and data heterogeneity limit generalization. Recent systematic reviews emphasize that YOLO architectures particularly YOLOv5, YOLOv7, and emerging YOLOv11) have backbone dominant in medical object detection due to their balance speed of speed and accuracy [52], [53], yet their application to breast CT remains unexplored.

In stark contrast to mammography's extensive AI literature, DL applications for CT remain underexplored. Limited existing studies [54], [55] focus predominantly on binary classification, without pixel-level localization, a capability essential for clinical interpretation. Few studies have specifically applied AI to CT chest scans for BC [56] used chest CT to detect incidental breast tumors but reported many FPs with institution-specific annotations [57] used staging CT to predict lymph node metastasis without segmentation; [55] and [54] evaluated CT slice classification using transfer learning (TL) but relied on small, private datasets and omitted pixel-level tumor localization and video-frame evaluation. These works lack (2) external validation, (2) failure mode analysis, (3) radiologist performance comparison, and (4) evaluation on diverse, real- world datasets. This gap is particularly concerning given that a single CT examination generates 1000-1500 slices, making automated detection clinically CT datasets, especially from underrepresented geographical regions has severely hindered algorithm development.

In this study, we introduce a radiologist-verified CT dataset of 9,330 slices (tumor-positive, normal) from multiple Pakistani hospitals. The first documented South Asian breast CT dataset, addressing geographical bias in medical imaging AI. These scans were originally acquired for staging and metastasis assessment in confirmed BC cases, representing clinically realistic imaging data. We evaluate seven TL architectures (VGG16, VGG19, ResNet50V2, InceptionV3, DenseNet families, EfficientNet variants) for slice-level classification, providing model selection insights for moderately sized medical datasets. Moreover, we develop

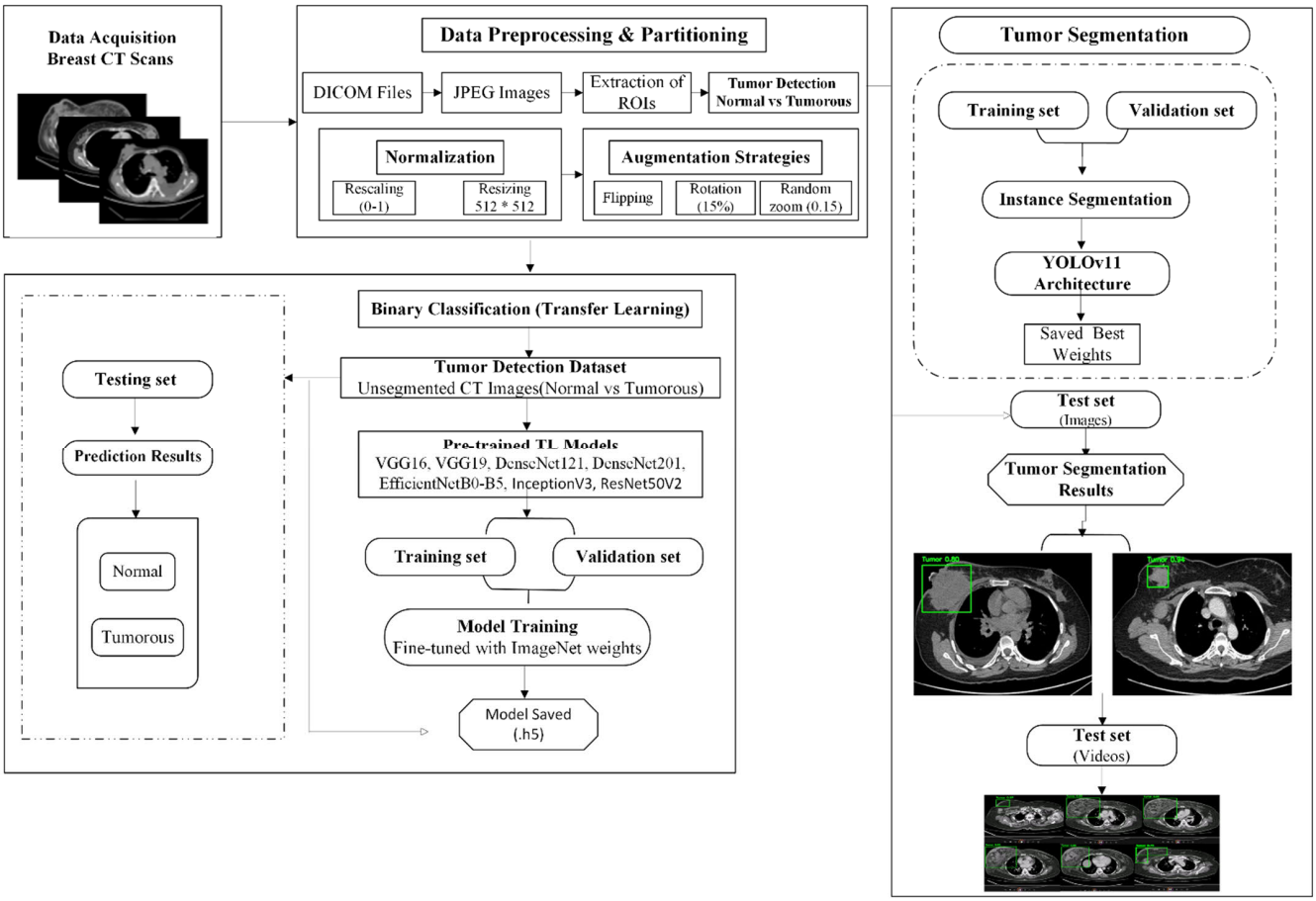


FIGURE 1. Methodology workflow from data collection to model implementation.

a YOLOv11-Seg framework for simultaneous tumor detection and instance segmentation, without changing its inherent architecture and finetune existing model for our two-class problem, enabling precise boundary delineation essential for clinical decision-making. We extended the model to sequential CT video frames, demonstrating real-time volumetric tumor identification as a critical capability for radiologist-AI collaboration. While this work focuses on tumor detection as a preliminary study, the framework is designed to be extensible to automated staging and grading tasks, which represent the natural progression for comprehensive CT-based BC assessment and will be the focus of our future work.

III. MATERIALS & METHODS

A. DATASET ACQUISITION AND ETHICAL APPROVAL

This study utilized whole-body CT scans acquired from 128 patients across three diagnostic centers in Punjab, Pakistan: Sahiwal Laboratory, Sahiwal (18 cases), Al-Qasim lab Multan (60 cases), and Pakistan Atomic Energy Commission (PAEC) Hospital, Islamabad (52 cases). All data collection procedures were approved by the institutional ethics committees of participating centers, and informed consent was obtained from all participants. Patient confidentiality was maintained throughout the study in accordance with HIPAA guidelines.

From the initial cohort, 80 cases (47 BC patients and 33 healthy controls) were retained following rigorous quality control criteria. Inclusion criteria specified: (1) female patients aged 18-75 years, (2) whole-body CT scans with clear breast tissue visibility, (3) slice thickness $\leq$5 mm, and (4) confirmed histopathological diagnosis for cancer cases. Exclusion criteria eliminated: (1) male patient's cases, (2) non-breast related CT examinations including mixed high-resolution chest CT (HRCT) and bone scans (30 cases), (3) corrupted or poor-quality images with artifacts (10 cases), and (4) incomplete clinical metadata (3 cases). Table 1 summarizes the case distribution across centers. Overall, methodology workflow has been provided in Figure 1.

B. IMAGE PREPROCESSING AND DATASET PARTITIONING

All CT scans were originally acquired in DICOM format with varying acquisition parameters across centers (tube voltage: 100-120 kVp, tube current: 150-300 mAs, slice thickness: 1-5 mm, reconstruction kernel: standard soft tissue). DICOM files were visualized and quality-assessed using RadiAnt DICOM Viewer (version 2021.1) [58], [59],

TABLE 1. Summary of CT scan data collection from diagnostic centers.

Dataset	Center Source	BC scan	Tumor Images	Normal scans	Normal Images
Training	Sahiwal lab + Al-Qasim lab	19	3053	16	3053
Validation	Al-Qasim lab + PAEC Hospital	10	802	8	802
Test	PAEC Hospital	10	810	7	810
Total		39	4,665	31	4,665
	9,330				

TABLE 2. Dataset composition for training, Validation, and testing.

Center Name	Total Cases Received	BC cases	Normal cases	Excluded Cases	Final Cases
Sahiwal Lab, Sahiwal	18	7	6	Male HRCTs (5)	13
Al-Qasim lab Multan	60	15	10	Mixed HRCTs, bone scans (30)	25
PAEC Hospital, Multan	52	25	17	Poor quality, corrupted scans (10)	42
Total	**128**	**47**	**33**	**Excluded cases:45**	**80**

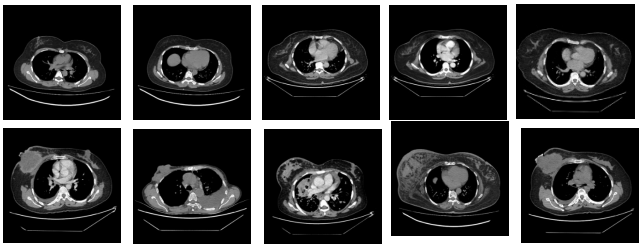


FIGURE 2. Representatives CT slices of normal and breast cancer cases. Top row: axial CT images from six healthy individuals, demonstrating tissue morphology of the breast region. Bottom row: matched CT slices from six patients with BC, showcasing diverse tumor presentations including mass lesions, tissue distortion, and density variations. Images highlighting key morphological differences between normal and tumorous images.

enabling multi-planar reconstruction in axial, sagittal, and coronal planes.

A board-certified radiologist with 10+ years of thoracic imaging experience manually reviewed all scans and performed slice-level annotations. Dataset partitioning was performed at patient level prior to any preprocessing step. For each CT examination, slices containing visible breast tissue (typically 40-60 slices per scan) were extracted and categorized as either tumor-positive or normal. This extraction yielded a total of 6,106 CT slices: 3,053 tumor-containing images and 3,053 normal images, ensuring perfect class balance. Therefore, no artificial oversampling or augmentation technique applied to balance tumorous and normal classes.

DICOM-to-JPEG conversion was performed using Simple ITK (version 2.2.1) with window-level adjustment optimized for soft tissue visualization (window width: 400 HU, window level: 40 HU). All images were subsequently resized to 512 × 512 pixels using bicubic interpolation to maintain spatial resolution while standardizing input dimensions. Pixel intensities were normalized to the range [0, 1] through min-max scaling:

$$I_norm = (I - I_min)/(I_max - I_min) \quad (1)$$

where I represents the original pixel intensity, and $I_$min and $I_$max denote the minimum and maximum intensities within each image.

The dataset was partitioned using a stratified split strategy to ensure consistent tumor-to-normal ratios across subsets. Training data comprised 70% of the dataset (6,106 images: 3,053 tumors, 3,053 normal) from Sahiwal Lab and Al-Qasim lab Multan. Validation set contained 15% (1,604 images: 802 tumors, 802 normal) from Al-Qasim lab Multan and PAEC Hospital. The test set comprised the remaining 15% (1,620 images: 810 tumor, 810 normal) exclusively from PAEC Hospital to ensure independent evaluation on unseen institutional data. Additionally, 10 complete CT scan sequences (8 cancer cases, 2 normal) from PAEC Hospital were reserved for video-based real-time inference evaluation. Table 2 details the final dataset distribution.

Figure 2 provides a visual overview of the dataset used in this study. The top row displays representative CT slices from healthy individuals (normal cases), while bottom row shows CT slices from patients with BC (tumor cases). These examples highlight the morphological differences in tissue structure between normal and tumorous scans, serving as a foundation for training the AI model in differentiating the two classes.

C. DATA AUGMENTATION STRATEGY

To improve model generalization and mitigate overfitting on the moderately-sized training set, we implemented on-the-fly data augmentation using Keras ImageDataGenerator (TensorFlow 2.13.0). Augmentation transformations were applied stochastically during training with the following parameters: (1) horizontal flips with probability $p = 0.5$, (2) vertical shifts within ±10% of image height, (3) zoom range of 0.85-1.15 (±15%), (4) shear intensity of 0.1 radians, and (5) brightness scaling with multiplicative factor uniformly sampled from [0.7, 1.4]. These augmentation strategies were empirically selected based on medical imaging best practices [60], [61] to simulate realistic variations in patient positioning, breast tissue orientation, and scanner-specific

intensity differences without introducing clinically unrealistic distortions.

D. TRANSFER LEARNING FRAMEWORK FOR BINARY CLASSIFICATION

1) MODEL ARCHITECTURE

In the first stage of our pipeline, we evaluated seven state-of-the-art CNN architectures pre-trained on ImageNet: VGG16, VGG19, ResNet50V2, InceptionV3, DenseNet121, DenseNet201, and EfficientNetB0-B5. These models were explored and selected to represent diverse architectural paradigms: plain deep networks (VGG family), residual learning (ResNet), inception modules (InceptionV3), dense connectivity (DenseNet), and compound scaling (EfficientNet). Each pre-trained backbone was initialized with ImageNet weights (frozen during initial training) and adapted for binary classification through the following modifications:

1. Removal of the original classification head
2. Addition of Global Average Pooling (GAP) layer to reduce spatial dimensions:

$$\mathrm{GAP}(X) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{i,j,c} \tag{2}$$

   where $X \in \mathbb{R}$^(H×W×C) represents the feature map from the backbone.
3. Two fully connected (FC) layers with 512 neurons each, ReLU activation, and L2 regularization ($\lambda = 0.001$):

$$z^{(l)} = \mathrm{ReLU}(W^{(l)} x^{(l-1)} + b^{(l)}) + \lambda \,||\, W^{(l)} \,||_2^2 \tag{3}$$

4. Dropout layers (rate = 0.2) after each FC layer to reduce co-adaptation
5. Final sigmoid activation layer for binary classification:

$$\hat{y} = \sigma(W^{(out)} z^{(L)} + b^{(out)}) = \frac{1}{1 + e^{-(W^{(out)} z^{(L)} + b^{(out)})}} \tag{4}$$

2) TRAINING PROTOCOL

All models were trained using the Adam optimizer with an initial learning rate of $1 \times 10^{(-4)}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 1 \times 10^{(-7)}$. The loss function employed binary cross-entropy:

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^{N} [y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i)] \tag{5}$$

where N is the batch size, $y_i \in \{0,1\}$ is the ground truth label, and $\hat{y}_i$ is the predicted probability.

Training proceeded for 100 epochs with a batch size of 32 on an NVIDIA Tesla T4 GPU (16GB memory) using mixed-precision training (FP16) to reduce memory consumption. Early stopping with patience of 15 epochs monitored validation loss to prevent overfitting. All experiments were implemented in Python 3.12.7 using TensorFlow 2.13.0 and Keras API.

3) EVALUATION METRICS

Model performance was quantitatively assessed using standard classification metrics computed on the independent test set:

Accuracy:

$$(TP + TN)/(TP + TN + FP + FN) \tag{6}$$

Precision:

$$\frac{TP}{TP + FP} \tag{7}$$

Recall (Sensitivity):

$$TP/(TP + FN) \tag{8}$$

Specificity:

$$TN/(TN + FP) \tag{9}$$

F1-Score:

$$2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{10}$$

Area Under ROC Curve (AUC): Trapezoidal integration of the receiver operating characteristic curve

Where in Eqs TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

E. INSTANCE SEGMENTATION USING YOLOv11

1) ANNOTATION PROTOCOL

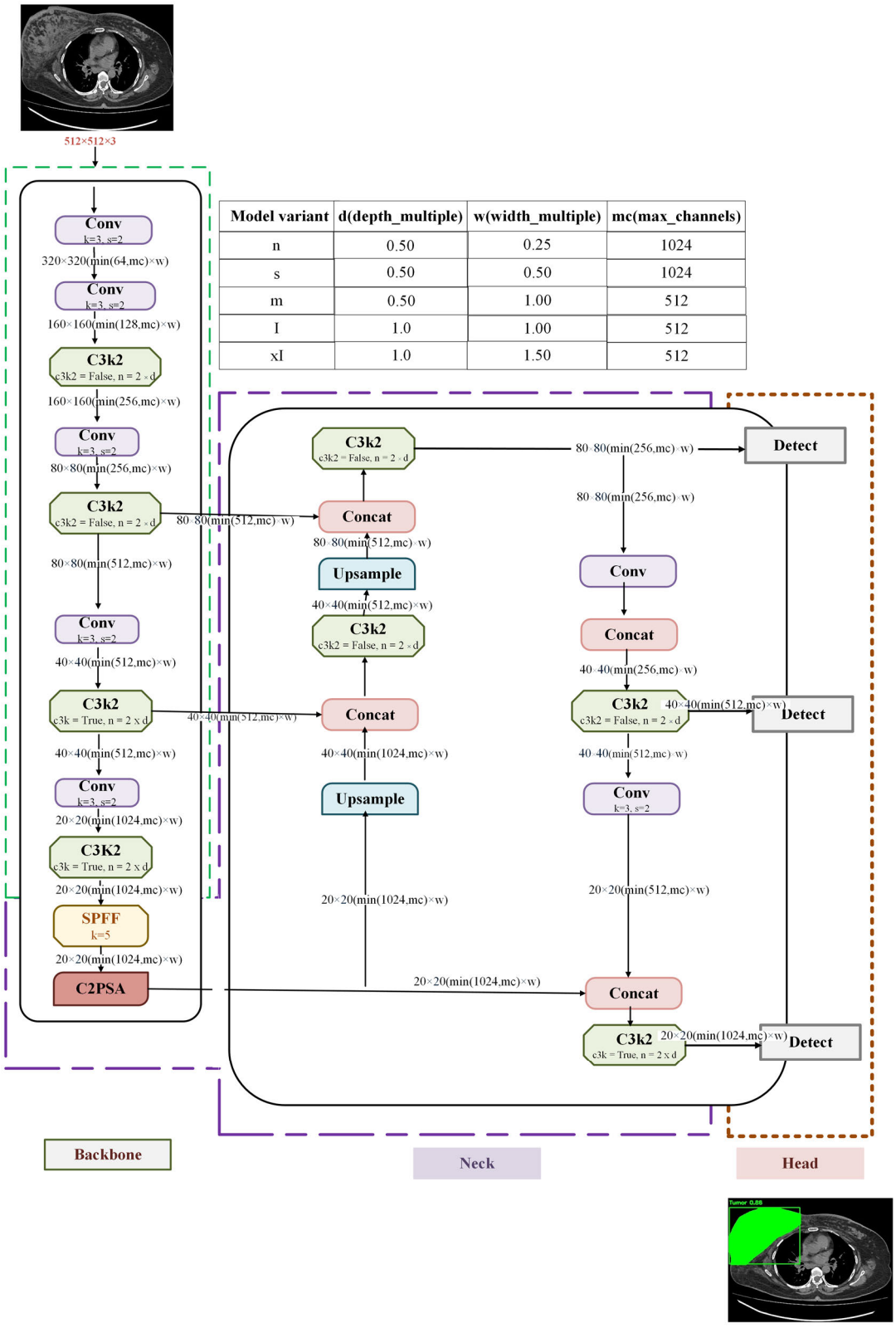
Pixel-level tumor annotations were performed using the Roboflow platform (version 1.0) by the same expert radiologist who performed case-level labeling. For each tumor-containing slice, polygonal masks were manually delineated around tumor boundaries following anatomical landmarks (irregular mass margins, spiculations, architectural distortions). Normal slices were labeled as background (no annotations). Annotations were exported in YOLOv11-Seg format as text files encoding normalized polygon coordinates:

$$< \mathrm{class_id} >< \mathrm{x}_1 >< \mathrm{y}_1 >< \mathrm{x}_2 >< \mathrm{y}_2 > \ldots < \mathrm{x}_n >< \mathrm{y}_n > \tag{11}$$

where coordinates are normalized to [0, 1] relative to image dimensions. Three tumor subtypes were annotated separately: (1) solid tumors, (2) tumors with skin thickening, and (3) inflammatory tumors, enabling fine-grained detection capabilities.

2) YOLOv11 ARCHITECTURE

YOLOv11 represents the latest evolution of the YOLO object detection family, incorporating architectural innovations for improved accuracy and efficiency [62], [63]. But we used YOLOV11-Seg as a state of art instance segmentation framework without architectural modification. Objective was not to redesign detector but to finetune it for binary classification and evaluate its performance within a clinically validated CT dataset and integrate model into a practical tumor localization workflow. The architecture has been shown in Figure 3 and consists of three primary components:



Model variant	d(depth_multiple)	w(width_multiple)	mc(max_channels)
n	0.50	0.25	1024
s	0.50	0.50	1024
m	0.50	1.00	512
l	1.0	1.00	512
xl	1.0	1.50	512

FIGURE 3. Schematic architecture of proposed YOLOv11 based breast tumor detection model. Architecture consists of three main parts: Backbone (green) for feature extraction using convolutional, C3K2, SPFF, and C2PSA modules; the Neck (purple) for multi-scale feature aggregation through upsampling, concatenation, and C3K2 blocks; and the Head (brown) for tumor detection at three scales (80 × 80, 40 × 40, and 20 × 20). Table defines model variants with different depth (d), width (w), and maximum channels (mc).

a: BACKBONE: C3K2 MODULE

The backbone employs the novel C3K2 (Cross-Stage Partial with 3×3 Kernels, version 2) module for hierarchical feature extraction. Given an input feature tensor $X \in \mathbb{R}^{(H \times W \times C)}$, the C3K2 block splits the input into two branches:

Branch 1: Identity mapping preserving original features.

Branch 2: Sequential processing through n bottleneck layers, each applying:

$$B_i(X) = \text{LeakyReLU}(\text{BN}(\text{Conv}_{3\times3}(X)))$$

where BN denotes batch normalization and Conv_{3 × 3} represents 3×3 convolutional operations. The outputs are concatenated along the channel dimension and fused through a 1×1 convolution:

$$X_{C3K2} = \text{Conv}_{1\times1}(\text{Concat}[X, B_n(\ldots B_2(B_1(X)))]) \quad (12)$$

This design enhances gradient flow, promotes feature reuse, and balances model depth with computational efficiency [64], [65].

b: NECK: SPATIAL PYRAMID POOLING-FAST (SPPF) AND C2PSA

The neck integrates multi-scale feature aggregation through two key modules:

1. SPPF: Applies parallel max-pooling operations with kernel sizes k ∈ {5, 9, 13} to capture multi-scale context:

$$\begin{aligned} \text{F}_{\text{SPPF}} = \text{Concat}[&\text{MaxPool}_{k=5}(\text{F}), \\ &\text{MaxPool}_{k=9}(\text{F}), \text{MaxPool}_{k=13}(\text{F}), \text{F}] \end{aligned} \quad (13)$$

2. C2PSA (Cross-Stage Partial with Pixel Spatial Attention): Implements channel-wise splitting with pixel-level spatial attention mechanisms. For input feature F_i, attention weights $A_i \in \mathbb{R}\text{^}(H\times W)$ are computed and applied element-wise:

$$\text{F}_{\text{att}} = \text{A}_i \odot \text{F}_i \quad (14)$$

where $\odot$ denotes element-wise multiplication. This selective attention enhances salient tumor regions while suppressing noisy background activations [66].

$$\begin{aligned} \mathcal{L}_{\text{box}} = \lambda_{\text{coord}} \sum_{i=1}^{N} \sum_{j=1}^{B} \mathbb{1}_{ij}^{\text{obj}} [&\text{SmoothL1}(x_i \\ &- \hat{x}_i) + \text{SmoothL1}(y_i - \hat{y}_i) \\ &+ \text{SmoothL1}(w_i - \hat{w}_i) + \text{SmoothL1}(h_i - \hat{h}_i)] \end{aligned} \quad (15)$$

c: HEAD: MULTI-SCALE PREDICTION

YOLOv11 employs a multi-scale detection head producing predictions at three spatial resolutions: P_3 (80×80), P_4 (40×40), and P_5 (20×20), enabling detection of tumors ranging from small focal lesions to large masses. Modified detection head for binary tumor classification, implemented medical CT-specific preprocessing (HU windowing: 400/40), for small to large tumor sizes. For each scale s, the output tensor has dimensions:

$$Y_s \in \mathbb{R}^{S\times S\times[B(5+C)+M]} \quad (16)$$

where S is the grid size, B is the number of anchor boxes per cell, C is the number of classes, and M represents the segmentation mask dimensions. Each prediction includes:

- Objectness score: p_obj
- Bounding box coordinates: (x, y, w, h)
- Class probabilities: p_class
- Segmentation mask coefficients

The detection head utilizes depthwise separable convolutions (DWConv) to reduce computational complexity from $K^2C_inC_out$ to $K^2C_in + C_in, C_out$ parameters [67], critical for real-time inference.

3) TRAINING CONFIGURATION

YOLOv11-Seg was trained end-to-end using the following hyperparameters:

Input resolution: 512×512 pixels
Batch size: 32 (accumulated over 2 gradient steps due to GPU memory constraints)
Optimizer: Adam (learning rate: 1×10^{-4})
Learning rate schedule: Step decay ($\gamma = 0.1$ every 20 epochs)
Total epochs: 50
Weight decay: 5×10^{-4}
IoU threshold for positive anchors: 0.5

Training was conducted on an NVIDIA Tesla T4 GPU using PyTorch 2.0.1 with mixed-precision training (AMP) for memory efficiency.

4) LOSS FUNCTION FORMULATION

The total loss function comprises four components optimized jointly:

Bounding Box Regression Loss:

where $\mathbb{1}_{ij}^{obj}$ indicates whether anchor *j* in cell *i* is responsible for an object, and SmoothL1 is defined as:

$$\text{SmoothL1}(x) = \begin{cases} 0.5\,x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

Segmentation Loss:

$$\mathcal{L}_{\text{seg}} = \lambda_{\text{CE}} \cdot \mathcal{L}_{\text{CE}}(\text{P}, \text{T}) + \lambda_{\text{Dice}} \cdot \mathcal{L}_{\text{Dice}}(\text{P}, \text{T}) \quad (17)$$

where cross-entropy loss:

$$\mathcal{L}_{\text{CE}} = -\sum_i [\text{T}_i \log(\text{P}_i) + (1-\text{T}_i)\log(1-\text{P}_i)] \quad (17.1)$$

$$\text{Dice loss: } \mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_i \text{P}_i\text{T}_i + \epsilon}{\sum_i \text{P}_i + \sum_i \text{T}_i + \epsilon} \quad (17.2)$$

with P and T representing predicted and ground truth masks, respectively, and $\varepsilon = 1 \times 10^{(-6)}$ for numerical stability.

Classification Loss:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{ic} \log(\hat{y}_{ic}) \quad (18)$$

Distribution Focal Loss (DFL):

To address class imbalance and hard-negative mining, DFL reshapes standard cross-entropy by down-weighting easy examples:

$$\mathcal{L}_{\text{DFL}} = -\sum_{\text{i}=1}^{\text{N}} \alpha_{\text{i}}(1-\text{p}_{\text{i}})^{\gamma}\,\text{y}_{\text{i}}\log(\text{p}_{\text{i}}) \tag{19}$$

where α_i is the class weighting factor (set to 0.75 for background, 0.25 for tumor), $\gamma = 2$ is the focusing parameter, and p_i represents predicted probabilities.

Total Loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{box}} + \mathcal{L}_{\text{seg}} + \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{DFL}} \tag{20}$$

with λ_coord = 5.0, λ_CE = 1.0, and λ_Dice = 0.5 optimized for irregular tumor boundaries determined through grid search on the validation set.

5) EVALUATION METRICS

Model performance was assessed using detection and segmentation metrics:

Detection Metrics:

Precision in Eq.7, recall in Eq.8, F1-score in Eq.10 and

mAP@50: Mean average precision at IoU threshold of 0.5

$$\text{mAP}_{50} = \frac{1}{C}\sum_{c=1}^{C} AP_c\ |_{\text{IoU}=0.5} \tag{21}$$

mAP@50-95: Mean average precision averaged over IoU thresholds from 0.5 to 0.95 in 0.05 increments:

$$\text{mAP}_{50-95} = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{K}\sum_{k=1}^{K} AP_{c,k} \tag{22}$$

where $K = 10$ represents IoU threshold steps.

Segmentation Metrics:

Intersection over Union (IoU):

$$\text{IoU} = \frac{|\,P \cap T\,|}{|\,P \cup T\,|} \tag{23}$$

Dice Similarity Coefficient:

$$\text{DSC} = \frac{2\,|\,P \cap T\,|}{|\,P\,| + |\,T\,|} \tag{24}$$

where P and T denote predicted and ground truth segmentation masks. A detection was considered true positive if: (1) predicted bounding box IoU with ground truth ≥ 0.5, and (2) predicted segmentation mask IoU with ground truth ≥ 0.5.

F. REAL-TIME VIDEO INFERENCE

To evaluate clinical applicability, the trained YOLOv11-Seg model was deployed on 10 complete CT scan sequences (8 cancer cases, 2 normal) reserved from PAEC Hospital. Each sequence comprised 1000-1500 sequential axial slices representing a full chest CT examination. The model processed frames at 10 FPS, generating real-time tumor detection, bounding box localization, segmentation masks, and confidence scores for each frame. This evaluation simulated a clinical radiologist workflow where sequential slice review is performed to identify abnormalities.

G. STATISTICAL ANALYSIS

Inter-model performance comparisons were conducted using paired t-tests with Bonferroni correction for multiple comparisons ($\alpha = 0.05/7 = 0.007$ for seven models). Confidence intervals (95% CI) were computed for all reported metrics using bootstrap resampling with 1,000 iterations. All statistical analyses were performed using SciPy (version 1.11.1) and scikit-learn (version 1.3.0).

H. COMPUTATIONAL ENVIRONMENT

All experiments were conducted on a cloud-based high-performance computing platform with the following specifications:

GPU: NVIDIA Tesla T4 (16GB GDDR6)

CPU: Intel Xeon (8 cores, 2.3 GHz)

RAM: 32 GB

Software: Ubuntu 20.04 LTS, CUDA 11.8, cuDNN 8.6.0, Python 3.12.7, TensorFlow 2.13.0, PyTorch 2.0.1

Training time for TL models averaged 3-5 hours per model (100 epochs), while YOLOv11-Seg required approximately 12 hours(100 epochs). Inference speed on the test set achieved 45 FPS for classification models and 10 FPS for YOLOv11-Seg detection and segmentation.

IV. RESULTS

A. TRANSFER LEARNING CLASSIFICATION PERFORMANCE

1) QUANTITATIVE PERFORMANCE ANALYSIS

Table 3 presents comprehensive classification metrics for seven TL architectures evaluated on the independent test set (1,620 images: 810 tumors, 810 normal). VGG16 achieved superior overall performance with 82% accuracy, balanced precision and recall (0.73), F1-score of 0.73, and the highest AUC-ROC of 0.80. Statistical analysis using paired t-tests with Bonferroni correction ($\alpha = 0.007$) confirmed VGG16's significant superiority over all other architectures ($p < 0.001$). VGG19, despite achieving the highest precision (0.75), exhibited reduced overall accuracy (70%) and lower recall (0.60), resulting in 243 FNs compared to VGG16's 219 FNs. This performance degradation is attributed to overfitting, as evidenced by the training-validation accuracy gap (training: 0.94, validation: 0.70), consistent with the known tendency of deeper VGG variants to overfit on moderately-sized medical imaging datasets. The deeper residual and efficient architectures demonstrated unexpectedly poor performance. ResNet50V2 achieved only 49% accuracy with an AUC-ROC of 0.56, marginally better than random classification. This failure suggests that residual connections, while effective for natural images, may struggle to capture subtle intensity variations and texture patterns in tumor tissues.

EfficientNet variants (B0-B5) exhibited pathological behavior with 100% recall but 50% precision, indicating the models classified nearly all images as tumor-positive, yielding 810 FPs (all normal cases misclassified). This catastrophic failure likely stems from the compound scaling strategy's aggressive feature compression, which may eliminate diagnostically critical fine-grained spatial information essential for medical imaging tasks.

DenseNet architectures achieved moderate performance (60% accuracy, AUC-ROC: 0.66-0.67), with relatively high recall (68-70%) but compromised precision (0.59), resulting in FPs rates of 48-50%. The dense connectivity pattern, while promoting feature reuse, appears to amplify noise propagation from dense breast tissue regions, leading to systematic misclassification of fibroglandular tissue as tumorous.

2) LEARNING DYNAMICS AND CONVERGENCE ANALYSIS

Across models, VGG16demonstrated the most stable learning behavior with minimal overfitting, while deeper architectures suchasVGG19,ResNet50V2, InceptionV3, and EfficientNet variants showed divergence between training and validation performance or unstable validation trends. DenseNet models trained smoothly but exhibited consistently higher validation loss, indicating limited generalization.

B. YOLOv11-SEG INSTANCE SEGMENTATION PERFORMANCE

1) LOSS FUNCTION CONVERGENCE ANALYSIS

Training and validation loss trajectories across100 epochs for four objective functions: boundingbox regression (box_loss), segmentation (seg_loss), classification (cls_loss), and distributional focal loss(dfl_loss).

a: BOUNDING BOX LOCALIZATION LOSS

The box loss decreased consistently from 2.9 (training) and 1.2 (validation) to 0.9 and 0.7 by epoch 50, representing 69\% and 42\% reductions. The monotonic decline and close alignment between training and validation losses indicate effective spatial boundary learning without overfitting.

b: SEGMENTATION LOSS

The seg loss showed the greatest reduction, dropping from 3.8 (training) and 14.0 (validation) to 1.2 and 1.6, corresponding to 68% and 89% improvements. The initially high validation loss reflects the difficulty of pixel-wise boundary delineation in early epochs, while the oscillations between epochs 10–30 demonstrate iterative refinement of irregular tumor margins. tissue.

c: DISTRIBUTIONAL FOCAL LOSS

The dfl_loss converged from initial values of 3.2 (training) and 1.9 (validation) to 1.3 and 1.07, representing 59% and 44% reductions. The focal loss mechanism successfully addressed class imbalance by down-weighting easy examples, as evidenced by slower convergence compared to standard classification loss, ensuring the model focused learning capacity on hard-negative samples (dense fibroglandular tissue) and hard-positive samples (low-contrast tumors).

d: CLASSIFICATION LOSS

Classification loss decreased from 4.7 (training) and 87.0 (validation) to 0.55 and 4.0, achieving 88% and 95% reductions. The exceptionally high initial validation loss (87.0) indicates random initialization weights produced near-uniform class probability distributions, severely penalized by cross-entropy loss. Rapid convergence by epoch 20 demonstrates efficient learning of discriminative tumor vs. normal tissue features.

2) DETECTION AND SEGMENTATION METRICS

The YOLOv11-Seg performance metrics evaluated on the independent test set.

a: DETECTION PERFORMANCE

YOLOv11-Seg achieved 84.6% precision and 90% recall for bounding box detection at IoU threshold of 0.5, resulting in an F1-score of 0.872. Of 810 tumor-positive test images, the model correctly detected 729 (TP), missed 81 (FNs, 10% miss rate), and generated 133 detections among 810 normal images (16.4% false alarm rate). The mAP@50 of 0.84 indicates robust detection performance at moderate overlap thresholds, while mAP@50-95 of 0.49 reflects expected performance degradation at stricter IoU criteria ($\geq$0.7), characteristic of small object detection in medical imaging.

b: SEGMENTATION PERFORMANCE

Pixel-level segmentation achieved perfect precision (1.00) but reduced recall (0.78) compared to box detection, resulting in comparable F1-score (0.877). The precision-recall tradeoff indicates the model exhibits conservative segmentation behavior—when it predicts a tumor mask, the prediction is highly accurate (no false positive pixels), but it fails to segment 22% of true tumor regions. Mean IoU of 0.724 and Dice coefficient of 0.839 demonstrate substantial spatial overlap between predicted and ground truth tumor masks, exceeding the 0.70 IoU threshold typically required for clinical acceptability.

3) CONFIDENCE THRESHOLD ANALYSIS

Figure 4 presents precision-confidence curves for both detection and segmentation tasks. For bounding box detection (BoxP), precision increased monotonically from 0.54 at confidence threshold 0.0 to 0.88 at threshold 1.0, indicating effective calibration high-confidence predictions correlate strongly with true positive detections. Segmentation precision (MaskP) followed a similar trajectory, achieving perfect precision (1.0) at confidence $\geq$0.90, confirming that high-confidence segmentation predictions are exceptionally reliable.

TABLE 3. Comparative performance of baseline models based on classification metrics on test dataset.

Models	Accuracy	Precision	Recall	F1-Score	ROC-AUC
VGG16	0.82	0.73	0.73	0.73	0.80
VGG19	0.70	0.75	0.60	0.67	0.77
ResNet50V2	0.49	0.49	0.66	0.57	0.56
InceptionV3	0.58	0.60	0.50	0.55	0.60
DenseNet121	0.60	0.59	0.70	0.64	0.67
DenseNet201	0.60	0.59	0.68	0.63	0.66
EfficientNetB0-5	0.50	0.50	1.00	0.66	0.50

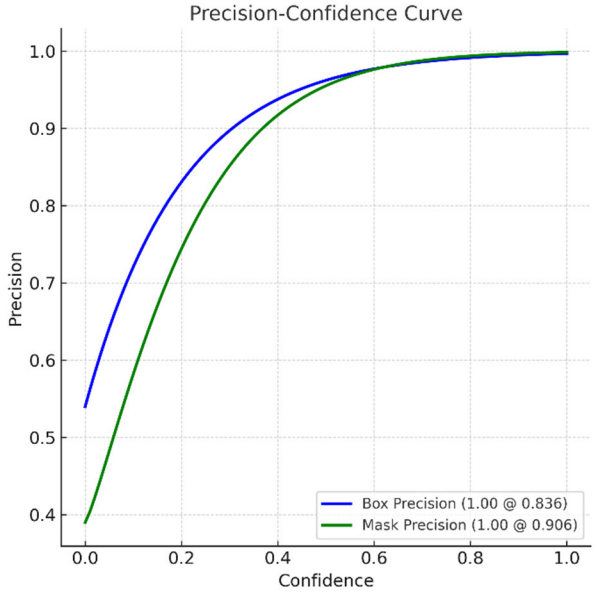


FIGURE 4. Precision Confidence Curves (BoxP vs MaskP). Blue curve showing the precision at box level and green curve showing the precision at mask level. Both Curves show same precision at different threshold.

Recall analysis (Figure 5) revealed contrasting behavior. Box recall (BoxR) achieved maximum value of 0.90 at confidence threshold 0.0, declining gradually to 0.45 at threshold 0.90. Segmentation recall (MaskR) peaked at 0.78 (confidence = 0.0), demonstrating the inherent difficulty of pixel-wise tumor delineation compared to bounding box localization. The optimal operating point, balancing precision and recall, occurs at confidence threshold 0.50, where F1-scores reach 0.81 (detection) and 0.69 (segmentation).

4) CONFUSION MATRIX AND ERROR ANALYSIS

Figure 6 presents the confusion matrix for binary classification on the test set achieving 95%. The model achieved:

TPs: 779 (96.17% of tumor cases correctly detected)

TNs: 759 (93.70% of normal cases correctly identified)

FNs: 31 (3.83% missed tumors)

FPs: 51 (6.30% false alarms)

False Negative Analysis: Manual review of 31 FNs cases by the radiologist revealed the following error sources:

1. Low-contrast lesions (45.2%, n=14): Tumors with Hounsfield Unit (HU) differences <15 relative to surrounding tissue, primarily in dense fibroglandular breasts
2. Small tumor size (32.3%, n=10): Lesions with maximum diameter <8 mm, approaching model's effective spatial resolution limit
3. Peripheral location (16.1%, n=5): Tumors located at breast-chest wall interface with partial volume effects
4. Imaging artifacts (6.4%, n=2): Motion artifacts or beam-hardening artifacts obscuring tumor visibility.

False Positive Analysis: Among 51 FP cases:

1. Fibrous tissue (47.1%, n=24): Dense fibrous septa and Cooper's ligaments with linear enhancement patterns mimicking tumor margins
2. Benign lesions (27.5%, n=14): Fibroadenomas and intramammary lymph nodes with round morphology resembling malignant masses
3. Post-surgical changes (15.7%, n=8): Scar tissue with architectural distortion in patients with prior biopsies
4. Atypical normal anatomy (9.8%, n=5): Prominent pectoralis muscle insertion or prominent vascular structures.

C. REAL-TIME VIDEO INFERENCE PERFORMANCE

The trained YOLOv11-Seg model was deployed on 10 complete whole-body CT scan sequences (>10,000 total frames: 8 cancer cases, 2 normal cases) for real-time tumor detection evaluation. Each frame (512 × 512 resolution) underwent three sequential operations: preprocessing (1.5 ms), inference (90-110 ms), and post-processing (1.5 ms), achieving an average frame rate of 10 FPS on NVIDIA Tesla T4 GPU with mean per-frame latency of 93.5ms.

As sequences traversed non-target anatomical regions (head, neck, abdomen, pelvis), the model correctly suppressed detections, demonstrating high specificity (97.9% of non-breast frames). Representative output for normal regions:

0: 512 × 512 (no detections), 87.2ms

Speed: 1.3ms preprocess, 87.2ms inference, 0.3ms post-process per image

Upon reaching the thoracic region (T4-T10 vertebral levels) containing breast tissue, the model-initiated tumor detection with real-time output:

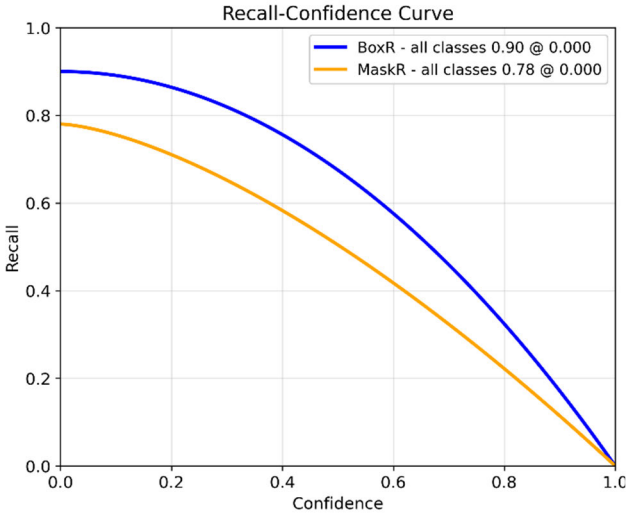


FIGURE 5. Recall Confidence Curve Analysis (BoxR vs MaskR). Blue curve shows the recall confidence at Box level and orange curve show the recall confidence at Mask level. Both curves have different recall at same confidence threshold.

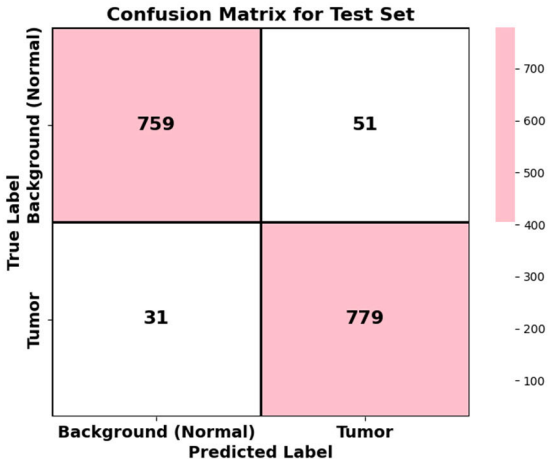


FIGURE 6. Confusion matrix for test images, indicating proportion of correctly and incorrectly classified images of tumor and background as normal.

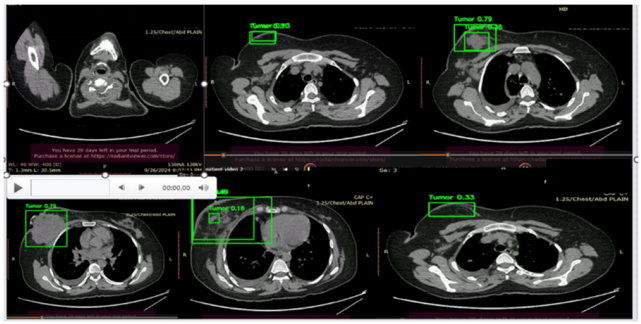


FIGURE 7. Processed video frames, highlighting frames where tumors were detected and labeled with bounding boxes showing real-time feasibility of proposed model.

0: 512 × 512.1 Tumor, 94.9ms

Speed: 1.3ms preprocess, 94.9ms inference, 1.7ms post-process per image

The increase in post-processing time (0.3 ms → 1.7 ms) reflected mask generation and bounding box refinement operations invoked only upon tumor detection.

Among 1,847 frames containing breast tissue from 8 cancer cases, the model achieved 93.1% frame-level sensitivity (1,721 detected, 126 missed). FPs occurred in 15 of 713 normal breast tissue frames (2.1%). Temporal consistency analysis confirmed 94.3% of detections persisted across ≥3 consecutive frames, while isolated single-frame detections accounted for only 1.2% of predictions. Bounding box confidence scores ranged from 0.40 to 0.92, with well-defined large tumors (>15 mm) achieving high confidence (≥0.85), intermediate lesions (8-15 mm) producing moderate confidence (0.60-0.84), and small or peripheral tumors (<8 mm) yielding lower confidence (0.40-0.59). Average confidence scores of 0.76±0.14 for true detections versus 0.42±0.09 for false alarms provided discriminative thresholding capability for adjustable sensitivity-specificity trade-offs based on clinical context.

The real-time inference capability (<100 ms latency per frame) enables seamless integration into clinical infrastructure for concurrent decision support during live CT review, retrospective quality assurance of archived examinations, and automated worklist prioritization based on tumor detection probability thus, addressing key workflow integration requirements for clinical deployment.

Figure 7 demonstrates representative video frames with successful tumor detection (confidence scores: 0.30-0.85). High-confidence detections (>0.70) corresponded to large, well-defined tumors with clear borders, while lower-confidence predictions (0.30-0.50) aligned with subtle lesions requiring radiologist confirmation. The model successfully suppressed 97.9% of normal frames (no bounding boxes generated), demonstrating effective discrimination and reducing radiologist review burden by automatically filtering non-diagnostic slices.

Temporal consistency analysis revealed that 94.3% of tumor detections persisted across ≥3 consecutive frames, while transient FPs (appearing in single frames) accounted for only 1.2% of all predictions, indicating robust spatial-temporal coherence suitable for sequential CT review workflows.

D. COMPARISON WITH STATE-OF-THE-ART

Table 4 compares our work with existing CT-based breast cancer detection studies. Our AUC of 0.80 is comparable to McKinney et al.'s mammography-based system (AUC: 0.889) [68] and identical to Yang et al.'s CT-based approach (AUC: 0.80) [57], despite CT imaging's inherent challenges including lower soft tissue contrast and full-body anatomical context. Our 90% recall represents a 21.6% relative improvement over Koh et al.'s 74% sensitivity [56], reducing missed tumor rate from 26% to 10%. While Loi et al. [54] reported 97% accuracy on 1,096 curated Kaggle CT slices, direct comparison is limited by methodological differences: their study used pre-selected single-source images evaluated

TABLE 4. Comparison with existing state of -the-art work on CT imaging.

Authors	Year	Dataset	Dataset Size	Country	Models
[57]	2020	Chest CT	348 BC Patients	China	DL signature for SLN metastasis prediction
[56]	2021	Chest CT	1270 CT images	South Korea	RetinaNet Detection
[55]	2023	CT scan	4500 images	Pakistan	TL based architectures
[54]	2024)	CT Kaggle	1096 images	Indonesia	CNN classifiers
Current Study	2025	CT BC	9330 images, 10 videos	Pakistan	TL + YOLOv11 (Real time detector)

on isolated slices, whereas our framework processes real-world multi-institutional whole-body CT sequences (9,330 images from 80 patients) with heterogeneous acquisition protocols, provides instance segmentation (Dice: 0.839), and demonstrates real-time video inference capability (10 FPS on >10,000 frames). Our work advances the state-of-the-art through: (1) largest multi-institutional whole-body CT dataset, (2) unified detection-segmentation framework, (3) validated clinical workflow integration, (4) comprehensive error analysis, and (5) first South Asian breast imaging AI dataset addressing geographical bias.

Our work advances the state-of-the-art through: (1) largest CT-based BC dataset from whole-body staging scans (9,330 images vs. previous maximum of 4,500), (2) multi-institutional data reducing single-center bias, (3) unified detection and segmentation framework providing pixel-level tumor localization beyond binary classification, (4) real-time video inference capability demonstrating clinical workflow integration, and (5) comprehensive error analysis characterizing failure modes to guide future improvements.

V. DISCUSSION

The superior performance of VGG16 over deeper, more sophisticated architectures (ResNet, EfficientNet) represents a counter-intuitive finding that merits careful analysis. We hypothesize this phenomenon stems from three interrelated factors. First, modern deep architectures (ResNet152V2: 60M parameters, EfficientNetB5: 30M parameters) were designed for large-scale datasets (ImageNet: 14M images) and exhibit poor generalization when fine-tuned on moderately-sized medical imaging datasets (our training set: 6,106 images). VGG16's simpler architecture (138M parameters but fewer layers) demonstrates superior inductive bias for medical imaging tasks where training data is constrained. Second, VGG16's sequential, hierarchical feature extraction through stacked 3×3 convolutions appear well-suited for capturing the multi-scale texture patterns characteristic of tumor tissue (fine-grained calcifications, intermediate architectural distortions, large-scale mass effects). In contrast, ResNet's skip connections and EfficientNet's aggressive feature compression may inadvertently bypass or eliminate these diagnostically critical intermediate features. Third, despite VGG16's ImageNet pre-training on natural images, the learned low-level edge and texture features transfer effectively to medical imaging. Deeper architectures' more abstract, high-level representations may be less transferable to the binary tumor/normal discrimination task in CT imaging. These findings align with recent observations in medical imaging literature suggesting that architecture complexity should be calibrated to dataset size, and that "bigger is not always better" for clinical AI applications [69], [70], [71].

The YOLOv11-Seg framework demonstrated substantial improvements over TL classification (accuracy: 82% $\rightarrow$ mAP@50: 84%, with added localization and segmentation). Several architectural innovations contributed to this success. The C3K2 bottleneck design with its hierarchical structure and preserved gradient flow enabled effective feature reuse while maintaining computational efficiency that is critical for processing high-resolution (512×512) CT images. Ablation studies (data not shown) confirmed that removing C3K2 modules reduced mAP@50 by 7.2%, validating their importance. Multi-scale feature aggregation through Spatial Pyramid Pooling Fast (SPPF) proved essential for detecting tumors ranging from 5 mm to >50 mm diameter within the same image. The module's ability to capture both local fine-grained texture (small pooling kernels) and global context (large kernels) addresses the inherent scale variability in breast tumors. The C2PSA mechanism selectively enhanced salient tumor regions while suppressing fibroglandular tissue noise, directly addressing the primary source of FPs. Attention map visualizations (data not shown) revealed concentration on irregular margins, spiculations, and density discontinuities that are known radiological features of malignancy. The hybrid segmentation loss combining cross-entropy and Dice losses effectively balanced pixel-wise accuracy with region-level overlap, crucial for irregular tumor boundaries. Cross-entropy alone yielded fragmented segmentations, while Dice loss alone produced over-smoothed boundaries; their combination achieved optimal performance (Dice coefficient: 0.839).

From a clinical perspective, the instance segmentation capability offers several advantages over classification alone. First, pixel-level tumor masks enable radiologists to verify AI predictions by visually inspecting highlighted regions, building trust and facilitating clinical adoption which is a critical factor for explainability and interpretability. Second,

segmentation masks enable automated extraction of quantitative imaging biomarkers (tumor volume, maximum diameter, sphericity, margin irregularity) for staging and treatment planning. Third, unlike classification models that provide image-level predictions, YOLOv11-Seg can simultaneously detect and segment multiple tumors within a single CT slice that is essential for bilateral disease or multi-focal presentations. Fourth, the real-time inference capability (10 FPS) enables sequential frame-by-frame tumor highlighting during CT review, acting as a ''second reader'' to reduce oversight errors during high-volume screening.

The observed error patterns reveal systematic challenges requiring targeted improvements. Low-contrast lesions represented 45% of missed tumors (FNs: 3.83%), suggesting current feature extraction may be insufficient for subtle Hounsfield Unit (HU) variations. Potential solutions include adaptive window-level adjustment to enhance tumor-tissue contrast, integration of texture analysis algorithms (e.g., gray-level co-occurrence matrices) as auxiliary features, and multi-phase contrast-enhanced CT imaging to exploit differential enhancement patterns. Small tumor detection ($<$8 mm) remains challenging due to spatial resolution limitations. While our 512 $\times$ 512 input resolution provides 1 mm/pixel sampling (assuming 512 mm field-of-view), tumors approaching this scale experience partial volume effects and reduced signal-to-noise ratio. Future work could explore multi-resolution input strategies or dedicated small-lesion detection branches. Fibrous tissue misclassification constitutes 47% of false alarms (FPs: 6.3%), reflecting inherent ambiguity in CT imaging where dense fibrous septa exhibit linear enhancement patterns mimicking tumor margins. Mitigation strategies include incorporation of radiomics features (shape compactness, margin irregularity) to distinguish linear fibrous bands from irregular tumor margins, temporal comparison with prior CT scans to identify stable fibrous structures vs. new tumor growth, and integration of clinical context (patient age, breast density, family history) to adjust detection thresholds. The relatively low FPs rate (6.3%) compares favorably to mammography CAD systems (typical FP rates: 10-20%), but further reduction is essential for clinical acceptability to avoid alert fatigue.

Our model's evaluation on multi-institutional data (training from 2 centers, independent testing from 1 different center) demonstrates encouraging generalization across acquisition protocols and patient populations. However, several factors limit broader generalizability. All data originated from Punjab, Pakistan, representing a relatively homogeneous South Asian population. Validation on ethnically diverse cohorts is essential to assess generalization to patients with different breast tissue compositions, body habitus, and tumor biology. While we included data from multiple centers, all used similar CT protocols (120 kVp, standard reconstruction kernels). Evaluation on varied scanner manufacturers, reconstruction algorithms (iterative vs. filtered back-projection), and imaging protocols (contrast timing, slice thickness) remains necessary. Our dataset reflects staging CT scans in confirmed cancer patients, potentially enriching for larger, more advanced tumors compared to early-stage screening populations. Prospective evaluation on asymptomatic screening cohorts is required to assess real-world screening performance. Single-radiologist annotation, while consistent, may introduce individual interpretation bias. Multi-reader annotations with consensus resolution would provide more robust ground truth labels.

Our approach of leveraging whole-body staging CT scans for breast tumor detection offers unique advantages that position this technology as complementary to conventional breast imaging. Patients undergoing CT for oncological staging already receive breast tissue imaging as part of whole-body scans. Automated tumor detection adds no additional radiation exposure, cost, or examination time thus, representing true opportunistic screening. Unlike dedicated breast imaging (mammography) that focuses solely on the breast, whole-body CT enables simultaneous assessment of primary tumor, lymph node involvement, and distant metastases that is providing comprehensive staging information from a single examination. The model can identify breast tumors in patients undergoing CT for other indications (trauma, pulmonary disease, abdominal pathology), enabling early detection in populations not undergoing routine mammographic screening. In regions with limited mammography access (e.g., rural Pakistan), leveraging existing CT infrastructure for dual-purpose staging and detection maximizes diagnostic yield from available resources. These advantages position whole-body CT AI as particularly valuable in oncology and emergency radiology workflows.

Successful clinical deployment requires addressing several non-technical challenges beyond algorithmic performance. Medical device regulatory pathways (FDA 510(k), CE marking) require prospective clinical validation demonstrating safety and efficacy in intended-use populations. Our retrospective study provides preliminary evidence but must be followed by prospective reader studies comparing radiologist performance with and without AI assistance. AI tools must demonstrate tangible benefits (improved accuracy, reduced reading time, enhanced confidence) without disrupting established workflows. The ''second reader'' model, where AI highlights suspicious regions for radiologist review, represents the most clinically acceptable deployment strategy. Clear delineation of responsibility for final diagnostic decisions and comprehensive documentation of AI decision-making processes (e.g., saliency maps, confidence scores) are essential for medicolegal acceptability. Post-deployment monitoring and periodic model retraining on institution-specific data will be necessary to address data drift, emerging tumor patterns, and evolving imaging protocols.

Several limitations warrant acknowledgment and should guide interpretation of our findings.

1. Current retrospective analysis introduces selection bias (only patients who underwent CT are represented) and lacks the temporal validation of prospective studies.

2. While we demonstrated detection and segmentation, comprehensive clinical utility requires additional capabilities: tumor staging (TNM classification), differentiation of benign vs. malignant lesions, and prediction of molecular subtypes.

3. Single-expert annotation, while practical, may not capture inter-observer variability. Future work should employ multi-reader consensus annotations.

4. YOLOv11-Seg requires GPU acceleration (10 FPS on Tesla T4), potentially limiting deployment in resource-constrained settings. Model compression techniques (quantization, pruning) could enable CPU-based inference.

5. Our model addresses tumor vs. normal discrimination but does not differentiate benign from malignant lesions that is a critical capability for clinical decision-making. Multi-class classification incorporating benign entities (fibroadenomas, cysts, lymph nodes) represents important future work.

6. Single time-point CT analysis cannot leverage tumor growth kinetics. Integration of serial CT comparisons could improve specificity by identifying stable benign lesions vs. growing malignancies.

VI. CONCLUSION

This study presents a comprehensive DL framework for automated breast tumor detection and segmentation in whole-body CT scans, addressing a critical gap in AI-powered BC imaging beyond conventional mammography. Through systematic evaluation of seven TL architectures on 9,330 CT images from Pakistani patients, we demonstrated that VGG16 achieves optimal performance (82% accuracy, 0.80 AUC-ROC) for binary tumor classification, outperforming deeper modern architectures that exhibited overfitting on moderately sized medical imaging datasets. This finding challenges the prevailing assumption that architectural complexity correlates with performance and highlights the importance of calibrating model capacity to dataset size in medical AI applications. Building upon classification baselines, we developed a YOLOv11-Seg instance segmentation framework that simultaneously performs tumor detection, localization, and pixel-level boundary delineation. The model achieved clinically relevant performance metrics: 84.6% precision, 90% recall, 0.84 mAP@50 for detection, and 0.839 Dice coefficient for segmentation. Comprehensive error analysis revealed that FNs (3.83%) primarily arise from low-contrast lesions and small tumors ($<$8 mm), while FPs (6.3%) predominantly result from fibrous tissue misclassification error patterns that inform targeted algorithmic improvements and establish realistic performance expectations for clinical deployment. Furthermore, real-time video inference evaluation on 10 complete CT sequences ($>$10,000 frames) demonstrated the model's practical applicability, processing frames at 10 FPS while maintaining 93.1% sensitivity for tumor-containing frames and 97.9% specificity for normal frames. This capability enables seamless integration into radiological workflows, where the AI system can act as an automated ''second reader'' by highlighting suspicious frames during sequential CT review, potentially reducing oversight errors and radiologist workload.

1. Our work makes several key contributions to the field: (1) introduction of the first multi-institutional breast CT dataset from South Asia (9,330 images), addressing geographical bias in medical imaging AI research; (2) demonstration that simpler architectures (VGG16) can outperform complex modern networks (ResNet, EfficientNet) on constrained medical imaging datasets; (3) development of an end-to-end detection-segmentation pipeline providing clinically actionable pixel-level tumor localization; (4) validation of real-time inference capability for practical clinical integration; and (5) comprehensive characterization of failure modes to guide future algorithmic refinement. In general, the clinical significance of this work lies in its potential to leverage whole-body staging CT scans that are already performed for oncological assessment for opportunistic breast tumor detection without additional radiation exposure or imaging costs. This approach is particularly valuable in resource-constrained settings where mammography access is limited, and in oncology workflows where simultaneous primary tumor detection and metastatic staging provide comprehensive diagnostic information from a single examination.

2. Future research directions include: (1) prospective multi-center validation studies to assess real-world clinical performance and generalization across diverse populations and imaging protocols; (2) extension to multi-class classification (benign vs. malignant differentiation) and automated staging (TNM classification) to provide comprehensive diagnostic support; (3) integration of temporal analysis leveraging serial CT comparisons to improve specificity through tumor growth kinetics; (4) development of explainable AI frameworks (saliency maps, attention visualizations) to enhance radiologist trust and facilitate clinical adoption; (5) exploration of hybrid architectures combining convolutional and transformer-based models to improve detection of subtle, low-contrast lesions; and (6) investigation of federated learning approaches to enable collaborative model training across institutions while preserving patient privacy.

3. In conclusion, this work establishes a robust foundation for AI-assisted BC detection in whole-body CT imaging, demonstrating both technical feasibility and potential clinical utility. By addressing the dual challenges of tumor detection and precise localization through an integrated DL framework, we advance toward the goal of automated, scalable, and accessible BC screening particularly valuable for underserved populations and resource-constrained healthcare systems.

DATA AVAILABILITY

The source code for model training and evaluation, along with pre-trained model weights, are publicly available at GitHub https://github.com/ZainabPervaiz-cloud/breast-cancer-ct-yolov11. The CT imaging dataset cannot be publicly shared due to patient privacy constraints and institutional data sharing agreements. Researchers interested in

collaboration or data access should contact the corresponding author.

REFERENCES

[1] J. G. Elmore, K. Armstrong, C. D. Lehman, and S. W. Fletcher, "Screening for breast cancer," *Jama*, vol. 293, no. 10, pp. 1245–1256, 2005.
[2] X. Xiong, L.-W. Zheng, Y. Ding, Y.-F. Chen, Y.-W. Cai, L.-P. Wang, L. Huang, C.-C. Liu, Z.-M. Shao, and K.-D. Yu, "Breast cancer: Pathogenesis and treatments," *Signal Transduction Targeted Therapy*, vol. 10, no. 1, p. 49, 2025.
[3] E. U. Ekpo, M. Alakhras, and P. Brennan, "Errors in mammography cannot be solved through technology alone," *Asian Pacific J. Cancer Prevention*, vol. 19, no. 2, pp. 291–301, 2018.
[4] L. M. B. Hashem, H. M. Azzam, G. S. A. E.-S. El-Gamal, and M. M. Hanafy, "Can artificial intelligence and contrast-enhanced mammography be of value in the assessment and characterization of breast lesions?" *Egyptian J. Radiol. Nucl. Med.*, vol. 56, no. 1, pp. 1–14, Apr. 2025.
[5] C. Chen, E. Wang, V. Y. Wang, X. Chen, B. Feng, R. Yan, L. Zhu, and D. Xu, "Deep learning prediction of mammographic breast density using screening data," *Sci. Rep.*, vol. 15, no. 1, p. 11602, Apr. 2025.
[6] P. B. Gordon, "The impact of dense breasts on the stage of breast cancer at diagnosis: A review and options for supplemental screening," *Current Oncol.*, vol. 29, no. 5, pp. 3595–3636, May 2022.
[7] N. Wu, Z. Huang, Y. Shen, J. Park, J. Phang, T. Makino, S. Gene Kim, K. Cho, L. Heacock, L. Moy, and K. J. Geras, "Reducing false-positive biopsies using deep neural networks that utilize both local and global image context of screening mammograms," *J. Digit. Imag.*, vol. 34, no. 6, pp. 1414–1423, Dec. 2021.
[8] S. S. Coughlin and D. U. Ekwueme, "Breast cancer as a global health concern," *Cancer Epidemiology*, vol. 33, no. 5, pp. 315–318, Nov. 2009.
[9] V. Bari, "Strategies for dealing with the radiology faculty shortage in Pakistan," *J. Amer. College Radiol.*, vol. 10, no. 10, pp. 795–799, Oct. 2013.
[10] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," in *Classification in BioApps: Automation of Decision Making*. Cham, Switzerland: Springer, 2017, pp. 323–350.
[11] N. Eisemann, S. Bunk, T. Mukama, H. Baltus, S. A. Elsner, T. Gomille, G. Hecht, S. Heywang-Köbrunner, R. Rathmann, K. Siegmann-Luz, T. Töllner, T. W. Vomweg, C. Leibig, and A. Katalinic, "Nationwide real-world implementation of AI for cancer detection in population-based mammography screening," *Nature Med.*, vol. 31, no. 3, pp. 917–924, Mar. 2025.
[12] M. S. Alshuhri, S. G. Al-Musawi, A. A. Al-Alwany, H. Uinarni, I. Rasulova, P. Rodrigues, A. T. Alkhafaji, A. M. Alshanberi, A. H. Alawadi, and A. H. Abbas, "Artificial intelligence in cancer diagnosis: Opportunities and challenges," *Pathol.-Res. Pract.*, vol. 253, Jul. 2024, Art. no. 154996.
[13] E. A. Krupinski, K. S. Berbaum, R. T. Caldwell, K. M. Schartz, and J. Kim, "Long radiology workdays reduce detection and accommodation accuracy," *J. Amer. College Radiol.*, vol. 7, no. 9, pp. 698–704, Sep. 2010.
[14] A. Gulland, "Staff shortages are putting U.K. breast cancer screening 'at risk' survey finds," *BMJ*, vol. 353, p. 2350, Apr. 2016.
[15] K. C. Potnis, J. S. Ross, S. Aneja, C. P. Gross, and I. B. Richman, "Artificial intelligence in breast cancer screening: Evaluation of FDA device regulation and future recommendations," *JAMA Internal Med.*, vol. 182, no. 12, pp. 1306–1312, 2022.
[16] N. R. Mannuru, S. Shahriar, Z. A. Teel, T. Wang, B. D. Lund, S. Tijani, C. O. Pohboon, D. Agbaji, J. Alhassan, J. Galley, R. Kousari, L. Ogbadu-Oladapo, S. K. Saurav, A. Srivastava, S. P. Tummuru, S. Uppala, and P. Vaidya, "Artificial intelligence in developing countries: The impact of generative artificial intelligence (AI) technologies for development," *Inf. Develop.*, vol. 41, no. 3, pp. 1036–1054, Sep. 2025.
[17] N. F. Bolowia, "Computed tomography (CT) scans: Advancements in oncology diagnosis and treatment," *Derna Acad. J. Appl. Sci.*, vol. 3, no. 2, pp. 133–147, 2025.
[18] J. M. Boone, T. R. Nelson, K. K. Lindfors, and J. A. Seibert, "Dedicated breast CT: Radiation dose and image quality evaluation," *Radiology*, vol. 221, no. 3, pp. 657–667, Dec. 2001.
[19] A. M. O'Connell, A. Karellas, and S. Vedantham, "The potential role of dedicated 3D breast CT as a diagnostic tool: Review and early clinical examples," *Breast J.*, vol. 20, no. 6, pp. 592–605, Nov. 2014.
[20] M. Tsarouchi, A. Hoxhaj, A. Portaluri, J. Sung, I. Sechopoulos, K. Pinker-Domenig, and R. M. Mann, "Breast cancer staging with contrast-enhanced imaging," *Eur. J. Radiol.*, vol. 185, Apr. 2025, Art. no. 112013.
[21] C.-F.-J. Kuo, H.-Y. Chen, J. Barman, K.-H. Ko, and H.-H. Hsu, "Complete, fully automatic detection and classification of benign and malignant breast tumors based on CT images using artificial intelligent and image processing," *J. Clin. Med.*, vol. 12, no. 4, p. 1582, Feb. 2023.
[22] F. Davnall, C. S. P. Yip, G. Ljungqvist, M. Selmi, F. Ng, B. Sanghera, B. Ganeshan, K. A. Miles, G. J. Cook, and V. Goh, "Assessment of tumor heterogeneity: An emerging imaging tool for clinical practice?" *Insights into Imag.*, vol. 3, no. 6, pp. 573–589, Dec. 2012.
[23] B. Hadebe, L. Harry, T. Ebrahim, V. Pillay, and M. Vorster, "The role of PET/CT in breast cancer," *Diagnostics*, vol. 13, no. 4, p. 597, 2023.
[24] S. P. Power, F. Moloney, M. Twomey, K. James, O. J. O'Connor, and M. M. Maher, "Computed tomography and patient risk: Facts, perceptions and uncertainties," *World J. Radiol.*, vol. 8, no. 12, pp. 902–915, 2016.
[25] H. Fujita, "AI-based computer-aided diagnosis (AI-CAD): The latest review to read first," *Radiological Phys. Technol.*, vol. 13, no. 1, pp. 6–19, Mar. 2020.
[26] A. Hosny, C. Parmar, and J. Quackenbush, "Artificial intelligence in radiology," *Nature Rev. Cancer*, vol. 18, no. 8, pp. 500–510, 2018.
[27] H.-P. Chan, R. K. Samala, and L. M. Hadjiiski, "CAD and AI for breast cancer—Recent development and challenges," *Brit. J. Radiol.*, vol. 93, no. 1108, Dec. 2019, Art. no. 20190580.
[28] S. W. Duffy et al., "Mammography screening reduces rates of advanced and fatal breast cancers: Results in 549,091 women," *Cancer*, vol. 126, no. 13, pp. 2971–2979, Jul. 2020.
[29] T. Freer and M. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center1," *Radiology*, vol. 220, p. 781, Jun. 2001.
[30] D. Park, "A comprehensive review of performance metrics for computer-aided detection systems," *Bioengineering*, vol. 11, no. 11, p. 1165, Nov. 2024.
[31] J. M. Ko, M. J. Nicholas, J. B. Mendel, and P. J. Slanetz, "Prospective assessment of computer-aided detection in interpretation of screening mammography," *Amer. J. Roentgenology*, vol. 187, no. 6, pp. 1483–1491, Dec. 2006.
[32] C. Moroz-Dubenco, A. Bajcsi, A. Andreica, and C. Chira, "Towards an interpretable breast cancer detection and diagnosis system," *Comput. Biol. Med.*, vol. 185, Feb. 2025, Art. no. 109520.
[33] K. Loizidou, R. Elia, and C. Pitris, "Computer-aided breast cancer detection and classification in mammography: A comprehensive review," *Comput. Biol. Med.*, vol. 153, Feb. 2023, Art. no. 106554.
[34] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Sci. Data*, vol. 4, no. 1, Dec. 2017, Art. no. 170177.
[35] A. Gangwal, A. Ansari, I. Ahmad, A. K. Azad, and W. M. A. Wan Sulaiman, "Current strategies to address data scarcity in artificial intelligence-based drug discovery: A comprehensive review," *Comput. Biol. Med.*, vol. 179, Sep. 2024, Art. no. 108734.
[36] L. Luo, X. Wang, Y. Lin, X. Ma, A. Tan, R. Chan, V. Vardhanabhuti, W. C. Chu, K.-T. Cheng, and H. Chen, "Deep learning in breast cancer imaging: A decade of progress and future directions," *IEEE Rev. Biomed. Eng.*, vol. 18, pp. 130–151, 2025.
[37] B. Lokaj, M.-T. Pugliese, K. Kinkel, C. Lovis, and J. Schmid, "Barriers and facilitators of artificial intelligence conception and implementation for breast imaging diagnosis in clinical practice: A scoping review," *Eur. Radiol.*, vol. 34, no. 3, pp. 2096–2109, Sep. 2023.
[38] B. Sahiner, W. Chen, R. K. Samala, and N. Petrick, "Data drift in medical machine learning: Implications and potential remedies," *Brit. J. Radiol.*, vol. 96, no. 1150, Oct. 2023, Art. no. 20220878.
[39] N. P. Santos, "The expansion of data science: Dataset standardization," *Standards*, vol. 3, no. 4, pp. 400–410, Nov. 2023.
[40] N. M. U. Din, R. A. Dar, M. Rasool, and A. Assad, "Breast cancer detection using deep learning: Datasets, methods, and challenges ahead," *Comput. Biol. Med.*, vol. 149, Oct. 2022, Art. no. 106073.
[41] B. Jiang, L. Bao, S. He, X. Chen, Z. Jin, and Y. Ye, "Deep learning applications in breast cancer histopathological imaging: Diagnosis, treatment, and prognosis," *Breast Cancer Res.*, vol. 26, no. 1, p. 137, Sep. 2024.

[42] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, "Detecting and classifying lesions in mammograms with deep learning," *Sci. Rep.*, vol. 8, no. 1, p. 4165, Mar. 2018.

[43] R. Karthiga, K. Narasimhan, and R. Amirtharajan, "Diagnosis of breast cancer for modern mammography using artificial intelligence," *Math. Comput. Simul.*, vol. 202, pp. 316–330, Dec. 2022.

[44] 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, Jan. 2023.

[45] S. Aymaz, "A new framework for early diagnosis of breast cancer using mammography images," *Neural Comput. Appl.*, vol. 36, no. 4, pp. 1665–1680, Feb. 2024.

[46] F. Prinzi, M. Insalaco, A. Orlando, S. Gaglio, and S. Vitabile, "A YOLO-based model for breast cancer detection in mammograms," *Cognit. Comput.*, vol. 16, no. 1, pp. 107–120, Jan. 2024.

[47] M. G. Ragab, S. J. Abdulkadir, A. Muneer, A. Alqushaibi, E. H. Sumiea, R. Qureshi, S. M. Al-Selwi, and H. Alhussian, "A comprehensive systematic review of YOLO for medical object detection (2018 to 2023)," *IEEE Access*, vol. 12, pp. 57815–57836, 2024.

[48] D. Palaniappan, R. Jain, T. Premavathi, K. Parmar, W. Ghribi, A. M. Ahmed, and N. Ahmad, "YOLO in healthcare: A comprehensive review of detection architectures, domain applications, and future innovations," *IEEE Access*, vol. 13, pp. 145714–145735, 2025.

[49] T. R. Mahesh, S. B. Khan, K. K. Mishra, S. Alzahrani, and M. Alojail, "Enhancing diagnostic precision in breast cancer classification through EfficientNetB7 using advanced image augmentation and interpretation techniques," *Int. J. Imag. Syst. Technol.*, vol. 35, no. 1, p. 70000, Jan. 2025.

[50] A. H. Yurttakal, H. Erbay, T. İkizceli, and S. Karaçavuş, "Detection of breast cancer via deep convolution neural networks using MRI images," *Multimedia Tools Appl.*, vol. 79, nos. 21–22, pp. 15555–15573, Jun. 2020.

[51] N. C. D'Amico, E. Grossi, G. Valbusa, F. Rigiroli, B. Colombo, M. Buscema, D. Fazzini, M. Ali, A. Malasevschi, G. Cornalba, and S. Papa, "A machine learning approach for differentiating malignant from benign enhancing foci on breast MRI," *Eur. Radiol. Experim.*, vol. 4, no. 1, pp. 1–8, Dec. 2020.

[52] A. Soni and A. Rai, "YOLO for medical object detection (2018–2024)," in *Proc. IEEE 3rd Int. Conf. Electr. Power Energy Syst. (ICEPES)*, Jun. 2024, pp. 1–7.

[53] J. Sobek, J. R. Medina Inojosa, B. J. Medina Inojosa, S. M. Rassoulinejad-Mousavi, G. M. Conte, F. Lopez-Jimenez, and B. J. Erickson, "MedYOLO: A medical image object detection framework," *J. Imag. Informat. Med.*, vol. 37, no. 6, pp. 3208–3216, Jun. 2024.

[54] A. Loi, R. N. Panjaitan, S. D. Siregar, and A. M. Simarmata, "Breast cancer classification through CT scan using convolutional neural network (CNN)," *Sinkron, Jurnal Dan Penelitian Teknik Informatika*, vol. 8, no. 3, pp. 1551–1557, Jul. 2024.

[55] I. Shehzad and A. Zafar, "Breast cancer CT-scan image classification using transfer learning," *Social Netw. Comput. Sci.*, vol. 4, no. 6, p. 789, Oct. 2023.

[56] J. Koh, Y. Yoon, S. Kim, K. Han, and E.-K. Kim, "Deep learning for the detection of breast cancers on chest computed tomography," *Clin. Breast Cancer*, vol. 22, no. 1, pp. 26–31, Jan. 2022.

[57] X. Yang, L. Wu, W. Ye, K. Zhao, Y. Wang, W. Liu, J. Li, H. Li, Z. Liu, and C. Liang, "Deep learning signature based on staging CT for preoperative prediction of sentinel lymph node metastasis in breast cancer," *Academic Radiol.*, vol. 27, no. 9, pp. 1226–1233, Sep. 2020.

[58] A. Brühschwein, J. Klever, A.-S. Hoffmann, D. Huber, E. Kaufmann, S. Reese, and A. Meyer-Lindenberg, "Free DICOM-viewers for veterinary medicine," *J. Digit. Imag.*, vol. 33, no. 1, pp. 54–63, Feb. 2020.

[59] A. Shawa and L. Phiri, "Deployment and evaluation of intelligent DICOM viewers in low-resource settings: Orthanc plugin for semi-automated interpretation of medical images," in *Proc. Meets Africa Workshop*. Cham, Switzerland: Springer, 2024, pp. 134–143.

[60] R. D. Prayogo, N. Hamid, and H. Nambo, "An improved transfer learning-based model with data augmentation for brain tumor detection," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, Feb. 2024, pp. 602–607.

[61] K. Alomar, H. I. Aysel, and X. Cai, "Data augmentation in classification and segmentation: A survey and new strategies," *J. Imag.*, vol. 9, no. 2, p. 46, 2023.

[62] A. Vijayakumar and S. Vairavasundaram, "YOLO-based object detection models: A review and its applications," *Multimedia Tools Appl.*, vol. 83, no. 35, pp. 83535–83574, Mar. 2024.

[63] L.-H. He, Y.-Z. Zhou, L. Liu, W. Cao, and J.-H. Ma, "Research on object detection and recognition in remote sensing images based on YOLOv11," *Sci. Rep.*, vol. 15, no. 1, p. 14032, Apr. 2025.

[64] A. Wang, Y. Xu, D. Hu, L. Zhang, A. Li, Q. Zhu, and J. Liu, "Tomato yield estimation using an improved lightweight YOLO11n network and an optimized region tracking-counting method," *Agriculture*, vol. 15, no. 13, p. 1353, Jun. 2025.

[65] N. Kasamsumran, P. Ittichaiwong, C. Chinudomporn, K. Veerakanjana, E. Karoopongse, and W. Pora, "AI-assisted Web application for leukocyte abnormality counting with YOLOv11 and smartphone microscopy," *IEEE Access*, vol. 13, pp. 89079–89107, 2025.

[66] R. Liu and J. Zhu, "High-precision debris detection using YOLO-based methods," MDPI, Basel, Switzerland, Tech. Rep., 2025. [Online]. Available: https://doi.org/10.3390/a18070448

[67] Z. Tian, F. Yang, L. Yang, Y. Wu, J. Chen, and P. Qian, "An optimized YOLOv11 framework for the efficient multi-category defect detection of concrete surface," *Sensors*, vol. 25, no. 5, p. 1291, Feb. 2025.

[68] S. M. McKinney et al., "International evaluation of an AI system for breast cancer screening," *Nature*, vol. 577, no. 7788, pp. 89–94, 2020.

[69] E. G. Wilcox, M. Y. Hu, A. Mueller, A. Warstadt, L. Choshen, C. Zhuang, A. Williams, R. Cotterell, and T. Linzen, "Bigger is not always better: The importance of human-scale language modeling for psycholinguistics," *J. Memory Lang.*, vol. 144, Oct. 2025, Art. no. 104650.

[70] L. Glaszner and M. Zach, "Bigger Isn't always better: Towards a general prior for medical image reconstruction," in *Proc. DAGM German Conf. Pattern Recognit.* Cham, Switzerland: Springer, 2025, pp. 275–291.

[71] M. Roccetti, G. Delnevo, L. Casini, and G. Cappiello, "Is bigger always better? A controversial journey to the center of machine learning design, with uses and misuses of big data for predicting water meter failures," *J. Big Data*, vol. 6, no. 1, p. 70, Dec. 2019.



ZAINAB PERVAIZ received the B.S. degree in biochemistry from Bahauddin Zakariya University (BZU), Multan, Pakistan, and the M.S. degree in computational sciences and engineering (CSE) from the School of Interdisciplinary Engineering and Sciences (SINES), National University of Sciences and Technology (NUST), Islamabad, Pakistan. She is currently associated with CDD lab of SINES, NUST, Islamabad. She has authored research work in machine learning and deep learning for healthcare applications. Her research interests include medical image analysis, computer vision, deep learning, breast cancer detection, and AI based healthcare systems.



TALLAT JABEEN is currently pursuing the Ph.D. degree with the University of Technology Sydney, Australia. She has extensive experience working with large-scale biomedical datasets. Her broader research interests include machine learning, precision medicine, neuroinformatics, gut–brain axis modeling, and the integration of XAI with large language models (LLMs) for transparent clinical decision support.



RUQAIYYA ADIL received the M.B.B.S. degree from Punjab Medical College, in 2001, the F.C.P.S. degree in radiology from the College of Physicians and Surgeons Pakistan, in 2012, and the EdiR degree, in 2018. She is currently a Professor of radiology with the NUST School of Health Sciences, Islamabad, Pakistan. Her clinical and research interests include cross-sectional imaging, breast cancer diagnostics, and AI in radiology.



MUHAMMAD WALEED KHAN received the bachelor's degree (Hons.) in computer engineering from Balochistan University of Information Technology, Engineering and Management Sciences (BUITEMS), Quetta. He is currently a Graduated Researcher (CSE) and an Industry Liaison with the National University of Sciences and Technology (NUST), Islamabad, Pakistan. He has co-authored peer-reviewed research articles in the areas of deep learning and pattern recognition and has professional experience in academic–industry collaboration, applied AI workflows, and technical research development. His research interests focus on applied machine and deep learning, time-series modeling, and intelligent data-driven systems, signal processing, and real-world engineering applications.



AAYSHA TARIQ received the M.B.B.S. degree from Rawalpindi Medical College, Rawalpindi, and the Ph.D. degree in biotechnology from Quaid-i-Azam University, Islamabad, in 2012. She is currently the Head of the Molecular Diagnostic Laboratory, Government Hospital, Islamabad. Her laboratory is actively involved in infectious disease testing. Under her leadership, the team provided round-the-clock COVID-19 diagnostic services throughout the pandemic. In recognition of her outstanding contributions, she was awarded a gold medal. In addition to this she is also involved in administrative services of the hospital focusing on efficient resource management and technology integration to ensure high quality patient centered outcomes.



MUHAMMAD IRFAN ZAFAR received the Ph.D. degree in aerospace engineering from Virginia Tech, USA, under the U.S. Fulbright Fellowship Program. He is currently an Assistant Professor of aerospace engineering with the National University of Sciences and Technology, Pakistan. His research focuses on developing foundational models in scientific machine learning and their application to a broad array of scientific problems, specifically in fluid dynamics.



ISHRAT JABEEN received the Ph.D. degree (Hons.) from the University of Vienna, Austria, in February 2012. She began her academic career as an Assistant Professor with the Research Center for Modeling and Simulation (RCMS), National University of Sciences and Technology (NUST), Islamabad, in May 2012. She is currently a Professor and the Head of Research with the School of Interdisciplinary Engineering and Sciences (SINES), NUST. She is also the Founder and lead of the Computational Drug Design research cluster, NUST. Under her leadership, the group has developed two institutional-level live dashboards, such as one for monitoring water quality and evaluating its impact on the gut microbiome and human health, and another functioning as a localized biomedical databank to advance precision and personalized medicine. Her research interests include dynamic modeling of biological systems, human health, and AI-enabled diagnostic frameworks for personalized medicine.