Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Heatmap-Guided Data-Centric YOLO Framework for Robust Breast Tumor Localization in Mammographic Images

**Zahra Samadi**
Guilan University of Medical Sciences
**Mohammad Norouzi**
millansamadi@gmail.com

Qa.c., Isalamic Azad University
**Seyed Omid Shahdi**
Qa.c., Isalamic Azad University
**Mona Haddad Zahmatkesh**
Guilan University of Medical Sciences

**Article**

**Keywords:** Breast cancer detection, Mammography, YOLO, Heatmap-guided learning, Data-centric augmentation, Deep learning

**Posted Date:** November 24th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11453814/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Heatmap-Guided Data-Centric YOLO Framework for Robust Breast Tumor Localization in Mammographic Images

Zahra Samadi[1], Mohammad Norouzi[2]*, Seyed Omid Shahdi[3], Mona Haddad Zahmatkesh[4]

1. GI Cancer Screening and Prevention Research Center, Guilan University of Medical Sciences, Rasht, Iran/ Department of Biomedical Engineering, Qa.c., Islamic Azad University, Qazvin, Iran.
2. Department of Electrical Engineering, Qa.c., Isalamic Azad University, Qazvin, Iran.
3. Department of Electrical Engineering, Qa.c., Isalamic Azad University, Qazvin, Iran.
4. GI Cancer Screening and Prevention Research Center, Guilan University of Medical Sciences, Rasht, Iran/ Department of Pharmaceutical Biotechnology–Nuclear Pharmacy, School of Pharmacy, Guilan University of Medical Sciences, Rasht, Iran.

*Corresponding author(s). E-mail(s): millansamadi@gmail.com, mh.norouzi@iau.ir;
Contributing authors: Z.samadi@iau.ir; shahdi@iau.ir; haddad@gums.ac.ir;

## Abstract

**Purpose**: Accurate localization of breast tumors in mammographic images is a critical yet challenging task due to low contrast, dense breast tissue and subtle lesion boundaries. These factors limit the performance of conventional computeraided detection systems and lightweight deep learning models. This study aimed to improve tumor localization accuracy and robustness through a data-centric, attention-guided detection strategy.
**Methods**: We proposed a heatmap-guided, data-centric object detection framework that leveraged neural attention to improve breast tumor localization in mammographic images. The framework integrated contrast-limited adaptive histogram equalization, bilateral denoising and neural heatmap generation derived from intermediate features of a lightweight YOLO-based detector. Unlike conventional augmentation strategies, the proposed approach used the generated heatmaps to guide lesion-aware data augmentation, enabling targeted enhancement of tumor-relevant regions and progressive construction of enriched training datasets. The detector was trained on high-resolution images using a consistent optimization strategy. Model performance was evaluated using standard object detection metrics, including mean average precision, precision, recall and loss convergence.
**Results**: The proposed heatmap-guided data augmentation significantly improved detection performance compared with the baseline dataset. Notable gains were observed in mean average precision across multiple thresholds, along with increased precision and recall. Training stability was improved and localization-related losses were consistently reduced, indicating more accurate and reliable tumor detection.
**Conclusion**: The results demonstrated that heatmap-guided, data-centric augmentation enhanced the accuracy, robustness and efficiency of lightweight YOLO-based models, providing a practical and interpretable solution for breast tumor localization in clinical mammography.

**Keywords**: Breast cancer detection; Mammography; YOLO; Heatmap-guided learning; Data-centric augmentation; Deep learning

## 1 Introduction

Breast cancer remains one of the leading causes of morbidity and mortality among women worldwide, representing a major public-health burden and a key driver of early diagnostic research [1–3]. The disease originates from uncontrolled cellular proliferation within breast tissue, where malignant cells can rapidly invade adjacent structures and metastasize to distant organs [4]. Despite significant advances in therapeutic options, the prognosis of breast cancer is critically dependent on early and accurate detection, as delayed diagnosis markedly increases treatment complexity, mortality risk and long-term clinical complications [5, 6]. Early detection has been shown to significantly improve survival rates and reduce the need for aggressive interventions, highlighting the need for robust, accessible and high-precision diagnostic technologies [7].
Mammography remains the gold-standard modality for breast-cancer screening due to its wide availability, noninvasive nature and ability to detect small abnormalities before they become symptomatic [8, 9]. However, accurate interpretation of mammographic images is an inherently difficult task influenced by multiple factors, including heterogeneous tissue density, low contrast, imaging artifacts, overlapping anatomical structures and the subtle morphological variations of tumor regions [10–12]. Moreover, false positives and false negatives remain prevalent due to the substantial visual similarity between benign and malignant findings and the inherent subjectivity associated with manual interpretation [9, 12, 13]. These challenges are exacerbated in datasets with “hard” characteristics, such as VinDr-Mammo, where variations in imaging quality, complex lesion patterns and nuanced diagnostic boundaries create additional difficulties even for expert radiologists [8, 14].
To address these limitations, computational systems based on image processing, machine learning (ML) and deep learning (DL) have gained considerable traction [15–17]. Classical image-processing techniques such as histogram-based enhancement, contour analysis and texture-feature extraction have demonstrated utility but often fail to capture the full complexity of lesion morphology, especially under variability in tumor shape, scale and location [17–19]. Conventional ML classifiers (e.g., SVM, KNN, XGBoost, Random Forest) overcome some weaknesses of hand-crafted features but remain constrained by their dependence on low-dimensional inputs derived from manual or semi-automated feature-engineering pipelines [17, 20]. Furthermore, most classification-based DL models, despite their strong performance in distinguishing normal vs. pathological images, cannot provide explicit tumor localization, which is essential for clinical decision-making, biopsy planning and treatment monitoring [18, 19, 21, 22].
In contrast, object-detection based deep neural networks offer a more comprehensive approach by simultaneously performing both classification and spatial localization. Among these, the YOLO (You Only Look Once) family of detectors has shown promising performance in medical imaging due to its real-time inference capability, end-to-end architecture and ability to process complex visual patterns across multiple scales [15, 23]. Recent studies applying YOLOv3, YOLOv4, YOLOv5 and YOLOv8 have demonstrated meaningful improvements in tumor recognition and bounding-box prediction [24, 25]. Nevertheless, the accuracy and robustness of these models still degrade when confronted with mammography datasets featuring low contrast, dense breast tissue, high noise or subtle lesion margins conditions especially predominant in hard datasets such as VinDr-Mammo [26–28].

To overcome these challenges, integrating advanced preprocessing and neural attention visualization has emerged as a promising strategy. Techniques such as Contrast Limited Adaptive Histogram Equalization (CLAHE) and bilateral filtering have proven effective in enhancing local contrast, reducing noise and increasing the visibility of subtle structures without oversmoothing critical edges [29–32]. In parallel, neural heatmaps derived from intermediate layers of deep networks enable the model to highlight regions of interest by capturing attention-weighted feature activations [33, 34]. When combined, such preprocessing and heatmap-based enhancements can guide object-detection models toward more precise localization and improved sensitivity to small or ambiguous lesions.

Building upon these insights, the present research proposes a novel framework for breast-tumor detection and localization that integrates neural heatmap fusion and CLAHE-based enhancement within an end-to-end object-detection pipeline. The system is designed to address several unresolved challenges: (1) improving YOLO-based sensitivity and specificity in hard mammography datasets, (2) reducing false positives and false negatives through enriched attention-guided feature extraction, (3) accelerating inference through model-optimization techniques such as quantization and computational simplification and (4) providing clinically interpretable localization outputs that support radiology decision-making. Using the intrinsic strengths of advanced DL architectures and complementary image-processing methods, this research aims to develop a robust, scalable and clinically relevant solution capable of performing high-accuracy tumor detection and precise spatial localization across diverse mammographic conditions.

Overall, this study contributes to the ongoing evolution of intelligent breast-cancer diagnostic systems by combining state-of-the-art object detection with enhanced preprocessing and heatmap-driven interpretability. The expected outcome is an automated framework with improved diagnostic reliability, reduced computational cost and greater clinical utility ultimately offering a promising pathway toward early detection, decision support and improved patient outcomes.

## 2 Methods

### 2.1 Overview

We propose a detection pipeline that combines image-level preprocessing and contrast enhancement with heatmap-guided YOLO-family object detectors to improve localization and classification of breast masses in digital mammography. The framework constructs enhanced training sets using a targeted augmentation strategy driven by heatmap analysis (lesion-aware augmentation), and evaluates several modern YOLO variants to identify the best trade-off between detection accuracy and computational efficiency for clinical deployment. Figure 1.

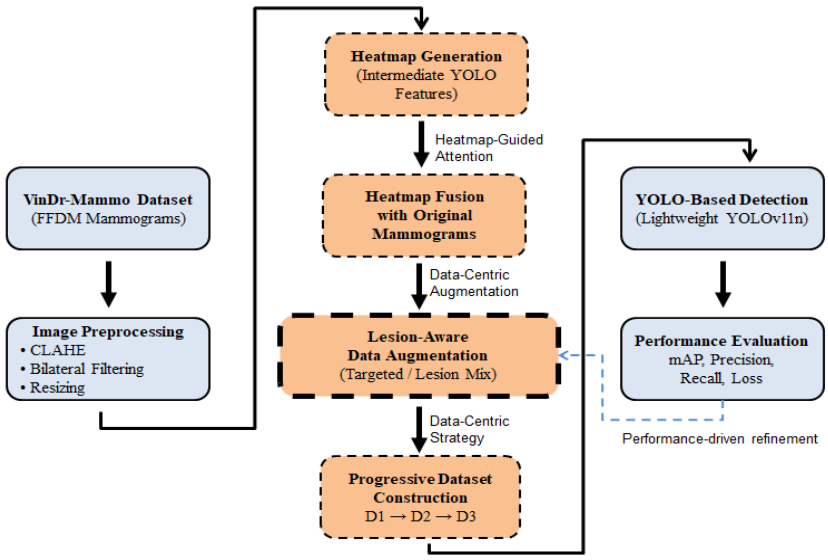



**Fig. 1.** Flowchart of Our Method: preprocessing, heatmap generation and fusion, lesion-aware augmentation (Lesion Mix) and YOLO-family detection and evaluation

### 2.2 Datasets

Experiments utilize publicly available mammography repositories and curated subsets, assembled into three distinct experimental datasets (D1–D3). The baseline dataset (D1) consists of original, annotated mammograms, with raw images converted from DICOM to PNG format and standardized for the pipeline. D2 represents a lesion-mix augmented dataset that enhances lesion diversity through synthetic augmentation techniques. D3 is an expanded training set derived from D2, with the final training volume approximately three times larger, used to explore the impact of additional targeted augmentations on model generalization. The VinDr-Mammo dataset, along with other widely used mammography collections, were employed for benchmarking and preprocessing validation.

The VinDr-Mammo dataset [35] is a large-scale, publicly available benchmark designed for computer-aided detection and diagnosis (CAD) in full-field digital mammography (FFDM). It comprises 5,000 screening examinations, each acquired in the standard four-

view protocol (left and right craniocaudal and mediolateral oblique views), resulting in a total of 20,000 high-resolution mammographic images. A key strength of VinDr-Mammo lies in the high-quality, multi-level annotation strategy. Each examination is independently reviewed by two board-certified radiologists, with discrepancies resolved through expert arbitration, ensuring strong inter-reader reliability. Lesion annotations are provided in the form of bounding boxes, making the dataset particularly suitable for object-detection and localization tasks. A visual representation of these datasets is provided in Figure 2.

### 2.3 Preprocessing pipeline

Preprocessing was designed to (1) reduce domain variability, (2) improve lesion conspicuity and (3) preserve geometric label consistency for object detection. The pipeline includes:

I. **Format conversion and resizing**. Raw DICOM images were converted to PNG and rescaled to a square input resolution appropriate for the detector. In the principal experiments images were resampled to 1024×1024; bounding-box coordinates were transformed with the same geometric mapping so that annotations remain consistent.

II. **Contrast enhancement.** We apply Contrast Limited Adaptive Histogram Equalization (CLAHE) [36] inside the segmented breast region to amplify local contrast while limiting noise amplification. CLAHE computes local histograms per tile, clips histogram counts at a clip limit and then applies histogram equalization. The resulting local intensity mapping is written as:

$$T(\mathrm{r}) = \left\lfloor (L-1)\sum_{k=0}^{r} p^{(k)} \right\rfloor \tag{1}$$

where $r$ is an input gray level, $L$ is the number of gray levels and $p(k)$ is the clipped probability density of level $k$ in the tile.

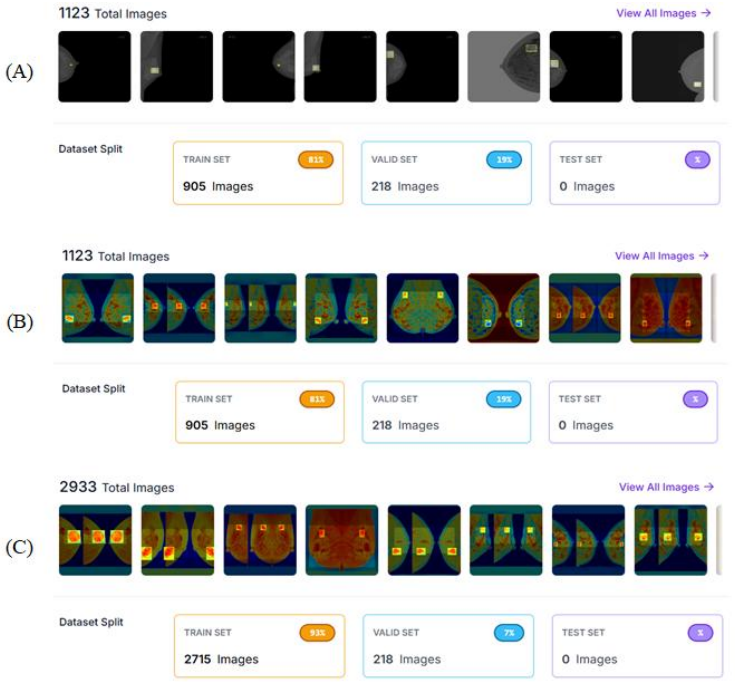



**Fig. 2.** Overview of the datasets utilized in this study: (A) D1: Baseline dataset, (B) D2: Preprocessed dataset and (C) D3: Expanded dataset with additional augmentation for enhanced model generalization

III. **Noise reduction**. To suppress high-frequency noise while preserving edges, we apply a bilateral filter [37]. The filtered intensity at pixel $x$ is:

$$I_f(x) = \frac{1}{W_p(x)} \sum_{y\in\Omega} I(y) g_s(\|x-y\|) g_r(|I(x)-I(y)|) \tag{2}$$

where $\mathrm{Wp(x)} = \sum_{y\in\Omega} I(y) g_s(\|x-y\|) g_r(|I(x)-I(y)|)$ is the normalization term, $g_s$ and $g_r$ are Gaussian kernels for spatial and range domains respectively and Ω is the filter window.

IV. **Color mapping for heatmap fusion.** For visualization and certain augmentation branches, a JET colormap (COLORMAP_JET) is applied to normalized heatmaps to produce colorized overlays used in augmentations.

V. **Bounding-box propagation**. All geometric operations (scaling, flips, cropping) are applied identically to the ground-truth bounding boxes so that annotation geometry remains correct after transforms.

## 2.4 Heatmap generation and candidate-region extraction

To bias training and augmentation toward diagnostically salient regions, mid-level feature maps were extracted from lightweight YOLO encoders to produce per-image saliency maps. Heatmaps are normalized and fused with the original intensity image using a weighted additive fusion:

$$I_{fused} = \alpha\, I + (1 - \alpha)\, H \tag{3}$$

where $I$ is the original intensity image, $H$ is the normalized heatmap (values scaled to the same dynamic range as $I$) and $\alpha$ is a fusion weight (empirically $\alpha = 0.7$ in our experiments). Candidate regions are then generated from $I_{fused}$ by thresholding and morphological refinement (connected-component filtering, small-region removal and optional dilation) to produce lesion-focused crops used both for lesion-aware augmentation and as auxiliary detector inputs.

## 2.5 Targeted data augmentation (Lesion-aware augmentation)

Using the candidate regions derived from heatmaps, we increase the representation of hard and under-represented lesions through targeted augmentations. Augmentation operators include controlled affine transforms (rotation, scaling), horizontal flipping, intensity perturbations, local contrast adjustments and the Lesion Mix operation, which inserts lesion patches into different breast backgrounds to synthetically increase lesion-background variability. Outputs of this pipeline produce datasets D2 and D3: D2 is Preprocessed dataset; D3 is an expanded set derived from repeated application of the targeted operations to reach roughly 3× the original training volume.

## 2.6 Object detection architecture

We benchmark several YOLO-family detectors and selected a lightweight YOLOv11n (Nano) backbone-head configuration as the main model due to its favorable balance of parameter count, FLOPs and inference speed. The YOLOv11n configuration used in experiments has approximately 2.59M parameters and 6.4 GFLOPs. Comparative experiments trained YOLOv8–YOLOv10 variants and alternate configurations on identical preprocessed inputs to assess sensitivity/specificity and inference tradeoffs. All detectors are trained to predict bounding boxes, objectness scores and lesion class labels.

## 2.7 Training procedure

Models were trained end-to-end on datasets D1–D3 with the following configuration:

• **Optimizer and schedule:** AdamW optimizer with initial learning rate lr = 0.002, weight decay = $5 \times 10^{-4}$, a cosine learning-rate scheduler and a short warm-up of three epochs.

• **Batching and epochs:** Training for 100 epochs with batch size 16.

• **Loss functions:** The composite YOLO loss comprises (i) bounding-box regression loss, (ii) classification loss and (iii) objectness loss. We use a Complete-IoU (CIoU) [38] style bounding-box loss and include a Distance Focal Loss (DFL) [39] term for improved localization. The CIoU loss is:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{4}$$

where IoU is the intersection-over-union between the predicted box $b$ and groundtruth $b^{gt}$, $\rho(0, 0)$ is the Euclidean distance between box centers, $c$ is the diagonal length of the smallest enclosing box covering $b$ and $b^{gt}$, $v$ measures aspect-ratio consistency and $\alpha = \frac{v}{1-IoU+v}$. The aspect-ratio term $v$ is:

$$v = \frac{4}{\pi^2}\left(\arctan\left(\frac{w^{gt}}{h^{gt}}\right) - \arctan\left(\frac{w}{h}\right)\right)^2 \tag{5}$$

where $w$, $h$ and $w^{gt}$, $h^{gt}$ are widths and heights of the predicted and ground-truth boxes respectively.

• **Distance Focal Loss**. For regression treated as a distribution over discrete locations, we use a DFL term. For a continuous label $y$ lying between discrete bin locations $y_i$ and $y_{i+1}$ with predicted probabilities $p_i$ and $p_{i+1}$, the DFL contribution is:

$$DFL = -[(y_{i+1} - y)\log(p_i) + (y - y_i)\log(p_{i+1})] \tag{6}$$

• **Data augmentation during training:** Standard affine augmentations, horizontal flips and the heatmap-guided targeted augmentations described above were applied on the fly.

• **Transfer learning:** Models were initialized from publicly available pretrained weights from corresponding YOLO families and fine-tuned on mammography data.

### 2.8 Evaluation metrics and protocol

Model performance was assessed using standard object-detection and localization metrics and per-class detection statistics.

• **Mean Average Precision (mAP).** We report mAP@0.50 (IoU threshold = 0.50) and COCO-style averaged mAP across IoU thresholds (0.50:0.95) [40]. The classaveraged mAP is computed as:

$$mAP = \frac{1}{N} \sum_{c=1}^{N} AP_c \quad (7)$$

where $AP_c$ is the average precision for class $c$. Average precision is computed as the area under the precision–recall curve; using a discrete summation it can be written as:

$$AP = \sum_k (P(k) - p(k-1))R(k) \quad (8)$$

where *P(k)* and *R(k)* are the interpolated precision and recall at operating point *k*.

• **Precision, recall, and F1.** Per-class precision P, recall R and F1-score are computed from detection counts:

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{TP+FN} \quad (10)$$

$$F1 = \frac{2PR}{P + R} \quad (11)$$

where *TP*, *FP* and *FN* denote true positives, false positives and false negatives respectively (match determined by IoU threshold and class agreement).

• **Training diagnostics**. Per-epoch monitoring of Box Loss, Class Loss and DFL Loss is used to assess training stability and to diagnose error modes.

Comparative experiments were run on D1–D3 to isolate the contribution of preprocessing and targeted augmentation. For each configuration we also report statistical dispersion of mAP across repeated training runs (variance and coefficient of variation) to quantify training stability.

### 2.9 Implementation details

All experiments were implemented in Python with PyTorch and the standard YOLO training toolchain. The primary reported configuration used 1024 × 1024 input resolution, AdamW optimizer and the hyperparameters listed above. Experimental logs, model checkpoints and evaluation scripts are archived and used to produce the quantitative results reported in Section 3.

## 3 Results

#### 3.1 Overview

The performance of the three datasets (D1, D2, D3) was evaluated using two primary metrics for object detection: Mean Average Precision (mAP) at IoU = 0.50 (mAP@0.50) and mAP averaged across multiple IoU thresholds (0.50 to 0.95), denoted as mAP@0.50:0.95. These metrics provide insights into both the accuracy and robustness of the models across different thresholds of overlap. Additionally, Precision, Recall and loss metrics were used to assess the detection performance in detail. The models were trained for 100 epochs, with the key findings summarized in the following subsections.

#### 3.2 Evaluation of mAP@0.50

We evaluate detection performance using mAP@0.50, which measures mean Average Precision at an Intersection-over-Union (IoU) threshold of 0.5. This metric is standard in object detection benchmarks and reflects localization accuracy under a moderate overlap criterion.

Figure 3 presents the mAP@0.50 trajectories over 100 training epochs for a fixed detector trained under three dataset setups (D1–D3). Training on D1, which employs minimal data augmentation, results in unstable optimization behavior, characterized by large oscillations and delayed convergence. In contrast, dataset setups D2 and D3, both incorporating progressively stronger augmentation strategies, lead to smoother learning dynamics and more consistent performance improvements throughout training.

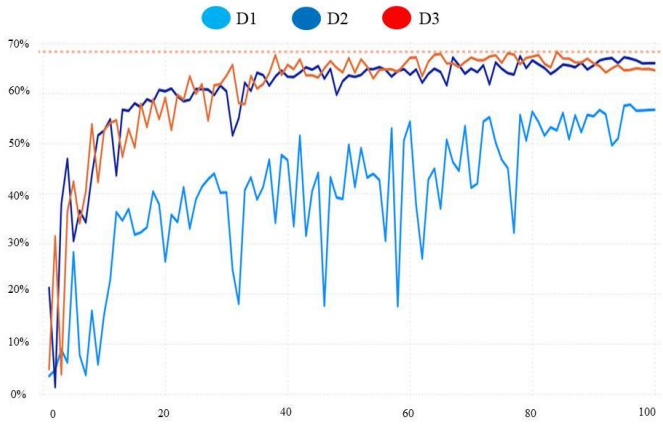



**Fig. 3.** mAP@0.50 over 100 epochs for a fixed detector trained on dataset setups D1, D2 and D3

### 3.2.1 Peak Performance Across Dataset Setups

Table 1 summarizes the maximum mAP@0.50 achieved when training on each dataset setup, along with the corresponding epoch. The detector trained on D3 achieves the highest peak performance, reaching 68.28% at epoch 84. Training on D2 yields a comparable maximum of 67.41% at epoch 78, while D1 attains a substantially lower peak of 57.74% at epoch 96.

**Table 1.** Maximum mAP@0.50 achieved under different dataset setups

| Dataset | Max mAP@0.50 | Epoch |
|---|---|---|
| **D1** | 57.74% | 96 |
| **D2** | 67.41% | 78 |
| **D3** | 68.28% | 84 |

Relative to D1, training on D2 and D3 improves peak mAP@0.50 by 9.67 and 10.54 percentage points, respectively. These gains indicate that the proposed augmentation and preprocessing strategies embedded in D2 and D3 substantially enhance detection performance, with D3 providing the most favorable data configuration.

### 3.2.2 Training Stability Analysis

To assess training stability, we compute the variance of mAP@0.50 over epochs 50–100. As shown in Table 2, the detector trained on D1 exhibits significantly higher variance compared to training on D2 or D3, indicating pronounced fluctuations during latestage optimization.

**Table 2.** Variance of mAP@0.50 over epochs50–100 under different dataset setups

| Dataset | Variance |
|---|---|
| **D1** | 0.00731 |
| **D2** | 0.000183 |
| **D3** | 0.000159 |

Both D2 and D3 reduce performance variance by more than an order of magnitude relative to D1, demonstrating that stronger data augmentation leads to markedly more stable convergence. Among the three, D3 achieves the lowest variance, indicating the most stable training behavior.

### 3.2.3 Coefficient of Variation

We further report the coefficient of variation (CV) of mAP@0.50 over epochs 50–100 to capture relative performance consistency. Table 3 shows that D3 yields the lowest CV (0.019), followed closely by D2 (0.021), whereas training on D1 results in substantially higher relative variability (0.176).

**Table 3.** Coefficient of Variation (CV) of mAP@0.50 over epochs 50–100

| Dataset | CV |
|---|---|
| **D1** | 0.176 |
| **D2** | 0.021 |
| **D3** | 0.019 |

The consistently lower variance and CV observed for D2 and D3 demonstrate that improved dataset construction not only boosts peak accuracy but also yields more reliable and stable training dynamics. Overall, D3 provides the most effective dataset setup in terms of both detection performance and robustness.

## 3.3 Evaluation of mAP@0.50:0.95

The mAP@0.50:0.95 metric evaluates detection performance by averaging precision over multiple IoU thresholds ranging from 0.50 to 0.95, providing a stringent assessment of localization robustness under varying overlap constraints. This metric is widely adopted to measure a detector's generalization ability beyond a single IoU operating point.
Figure 4 illustrates the evolution of mAP@0.50:0.95 over training epochs for models trained on datasets D1, D2, and D3. Across all epochs, the model trained on D3 consistently achieves higher mAP@0.50:0.95 values, indicating improved robustness to stricter localization requirements compared to training on D1 or D2.
Table 4 reports the maximum mAP@0.50:0.95 achieved by models trained on each dataset, along with the epoch at which the peak performance occurs.

**Table 4.** Maximum mAP@0.50:0.95 achieved by models trained on different datasets

| Training Dataset | Maximum mAP@0.50:0.95 | Epoch |
|---|---|---|
| **D1** | 32.80% | 95 |
| **D2** | 39.35% | 68 |
| **D3** | 40.82% | 38 |

Models trained on D3 achieve the highest mAP@0.50:0.95 of 40.82%, outperforming those trained on D2 (39.35%) and D1 (32.80%). Notably, the peak performance for D3 is reached substantially earlier (epoch 38) than for D2 (epoch 68) and D1 (epoch 95), suggesting that the data composition and augmentation strategy in D3 facilitate both improved generalization across IoU thresholds and faster convergence. Overall, these results highlight the importance of dataset design in enhancing detection robustness under strict localization criteria.

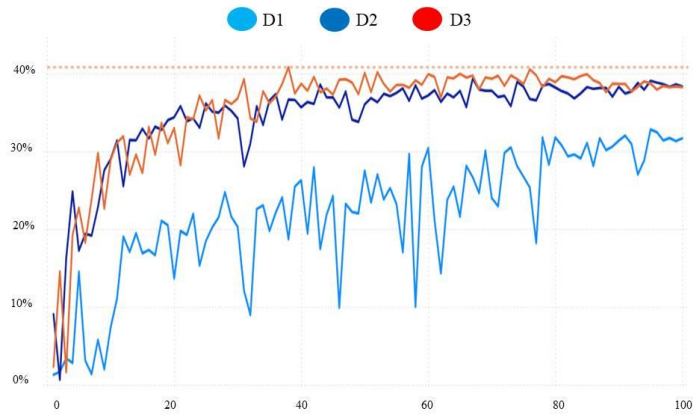



**Fig. 4.** mAP@0.50:0.95 over training epochs for models trained on datasets D1, D2 and D3

### 3.4 Precision and Recall Analysis

Precision and Recall are standard metrics for evaluating the correctness and sensitivity of object detection systems. We report both metrics across training epochs to analyze how performance varies across different datasets and to highlight the precision–sensitivity trade-off that is particularly relevant in medical imaging applications.

#### 3.4.1 Precision analysis

Figure 5 illustrates the Precision curves obtained on datasets D1, D2 and D3 over training epochs.
Among the evaluated datasets, D3 achieves the highest peak Precision of 76.41% at epoch 92. Dataset D2 follows closely with a maximum Precision of 75.42%, while D1 exhibits a substantially lower peak Precision of 62.73%. The higher Precision observed on D3 indicates that, for this dataset, the detector produces fewer false-positive predictions relative to true positives. This suggests that D3 provides visual characteristics or annotations that better support precise localization, leading to more reliable positive detections.

#### 3.4.2 Recall analysis

Figure 6 reports Recall performance on the same datasets throughout training.
Dataset D3 again yields the strongest performance, achieving a maximum Recall of 64.76%. D2 attains a comparable Recall of 63.72%, whereas D1 lags behind with a peak Recall of 54.38%. The higher Recall on D3 demonstrates improved sensitivity, indicating that a larger proportion of true lesions are successfully detected. In the context of breast cancer screening, such sensitivity is critical, as false negatives may lead to missed diagnoses. The comparatively lower Recall on D1 suggests that detection on this dataset is more challenging, potentially due to differences in image quality, lesion appearance or annotation variability.
Taken together, these results show that dataset characteristics substantially influence detection performance. D3 consistently enables both higher Precision and Recall, while D1 poses a more challenging detection scenario.

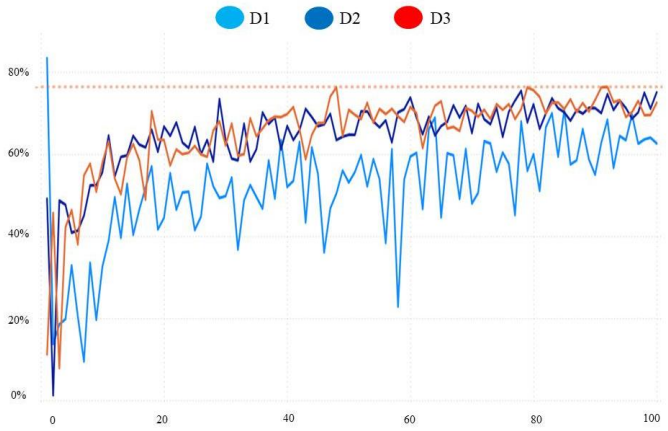



**Fig. 5.** Precision on datasets D1, D2 and D3 across training epochs

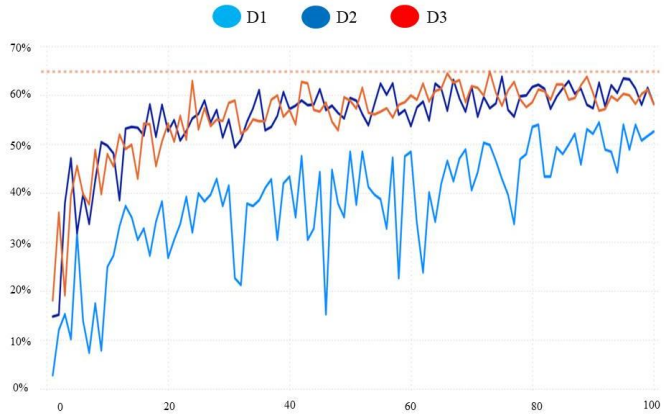



**Fig. 6.** Recall on datasets D1, D2 and D3 across training epochs

## 3.5 Error Analysis via Loss Functions

To further analyze detection behavior across datasets, we examine the training losses on the validation split, including Box Loss, Classification Loss and DFL. These losses provide complementary insights into localization accuracy, category discrimination and bounding-box regression quality, respectively.

*Overview of loss terms*: Box Loss evaluates the geometric alignment between predicted and ground-truth bounding boxes. Classification Loss, implemented via cross-entropy, measures the discrepancy between predicted and true class labels. DFL refines bounding-box regression by modeling localization as a discrete probability distribution, enabling more precise coordinate estimation.

### 3.5.1 Box Loss

Table 5 reports the Box Loss values measured on the validation set for datasets D1–D3.

**Table 5.** Box Loss on the validation set.

| Dataset | Box Loss |
|---|---|
| **D1** | 1.79 |
| **D2** | 1.55 |
| **D3** | 1.53 |

Dataset D3 yields the lowest Box Loss, closely followed by D2, while D1 exhibits substantially higher localization error. This indicates that bounding-box regression is more accurate and stable on D3, suggesting more consistent lesion boundaries or higher-quality annotations.

### 3.5.2 Classification Loss

Classification Loss values on the validation set are summarized in Table 6.

Dataset D3 achieves the lowest Classification Loss, indicating more accurate class predictions and reduced ambiguity between lesion categories. D2 also shows improved performance compared to D1, while the higher loss on D1 suggests greater class imbalance, visual ambiguity, or annotation noise.

#### 3.5.3 Distance Focal Loss

Table 7 presents the DFL values for each dataset.

**Table 6.** Classification Loss on the validation set

| Dataset | Class Loss |
|---|---|
| **D1** | 2.33 |
| **D2** | 1.60 |
| **D3** | 1.48 |

**Table 7.** DFL on the validation set

| Dataset | DFL Loss |
|---|---|
| **D1** | 1.69 |
| **D2** | 1.53 |
| **D3** | 1.46 |

Consistent with Box Loss trends, D3 exhibits the lowest DFL, reflecting more precise and stable localization of bounding-box coordinates. The higher DFL on D1 indicates increased uncertainty in regression targets, reinforcing the greater difficulty of this dataset.

Across all loss components, dataset D3 consistently enables lower validation errors, followed by D2, with D1 posing the most challenging detection scenario. These findings align with the observed improvements in Precision, Recall and mAP, and demonstrate that dataset characteristics, such as image quality, lesion visibility and annotation consistency, play a critical role in training stability and detection accuracy.

Overall, the loss-based error analysis corroborates the quantitative performance metrics and highlights the importance of robust preprocessing and augmentation strategies in improving both localization and classification reliability for breast tumor detection in digital mammography.

## 4 Discussion

This study examines the effectiveness of heatmap-guided preprocessing and lesion-aware data augmentation for improving object-detection based breast tumor localization in digital mammography. Through systematic experiments across progressively enriched dataset configurations (D1–D3), we demonstrate that data-centric design choices play a decisive role in improving detection accuracy, training stability and convergence behavior, even when using lightweight YOLO-family detectors.

### 4.1 Effect of Lesion-Aware Augmentation on Detection Performance

A key finding of this work is the substantial improvement in detection performance achieved through lesion-aware augmentation. Transitioning from the baseline dataset (D1) to the heatmap-guided augmented datasets (D2 and D3) results in gains exceeding 10 percentage points in mAP@0.50 and nearly 8 percentage points in mAP@0.50:0.95. These improvements indicate enhanced robustness not only under moderate overlap criteria but also under stricter localization thresholds.

Importantly, the observed gains cannot be attributed solely to increased data volume. Rather, the targeted augmentation strategy enriches lesion appearance diversity and lesion background interactions, enabling the detector to learn more discriminative and generalizable representations. Dataset D3 consistently achieves the highest peak performance while requiring fewer training epochs to converge, suggesting that heatmap-guided augmentation provides more informative supervisory signals early in training.

### 4.2 Training Stability and Optimization Behavior

Beyond peak accuracy, our analysis highlights notable improvements in training stability. Models trained on D2 and D3 exhibit more than an order of magnitude reduction in late-stage mAP variance compared to D1. This stability is reflected in the smoother learning curves and is further supported by lower validation Box Loss, Classification Loss and DFL.

The reduced Box Loss and DFL values observed for D3 indicate more precise and consistent bounding-box regression, particularly in cases involving ambiguous lesion boundaries or dense tissue. Similarly, lower Classification Loss suggests improved class separability, reducing confusion between visually similar benign and malignant findings. These results emphasize the importance of robust dataset construction in stabilizing optimization dynamics for medical object detection tasks.

#### 4.3 Role of Heatmap-Guided Attention and Interpretability

The integration of neural heatmaps into both preprocessing and data augmentation serves a dual purpose. First, heatmaps act as weak localization priors, biasing training toward diagnostically salient regions and mitigating the influence of irrelevant background structures. Second, they enhance interpretability by enabling qualitative inspection of model attention, thereby increasing transparency and clinical trustworthiness.

Unlike post-hoc explainability techniques, the proposed framework incorporates attention cues directly into the data generation process. This design encourages the model to internalize meaningful visual patterns throughout training, aligning with recent trends in attention-guided and interpretable medical AI [41–44].

### 4.4 Efficiency–Accuracy Trade-offs and Clinical Implications

An important practical implication of this study is that the reported performance gains are achieved using a lightweight YOLOv11n architecture with a modest parameter count and computational footprint. This finding demonstrates that architectural scale is not the sole determinant of performance in mammographic detection tasks; instead, data-centric enhancements can unlock the full potential of compact models.

From a clinical perspective, the simultaneous improvements in Precision and Recall observed on D3 are particularly relevant. Higher Recall reduces the likelihood of missed lesions, while improved Precision limits false-positive detections that may lead to unnecessary follow-up procedures. The availability of reliable bounding-box localization further supports downstream clinical workflows such as lesion review and biopsy planning.

### 4.5 Limitations and Future Directions

Despite the encouraging results, several limitations remain. First, while the proposed framework improves robustness on challenging dataset, cross-domain generalization across imaging devices, acquisition protocols and patient populations has not been fully explored. Future work may incorporate multi-institutional evaluation and domain adaptation strategies.

Second, the current study focuses on bounding-box level detection. Extending the framework to pixel-level segmentation or multi-lesion detection could further enhance clinical utility. Finally, prospective studies involving radiologist in the loop evaluation would be valuable to assess real-world usability, interpretability and diagnostic impact.

## 5 Conclusion

This study demonstrated that heatmap-guided preprocessing combined with lesion-aware data augmentation effectively improves breast tumor localization in digital mammography. By integrating neural attention information directly into the data construction process, the proposed framework consistently enhanced localization accuracy, precision and recall while reducing localization and classification losses. The results further showed improved training stability and faster convergence, indicating that performance gains were primarily driven by targeted enrichment of lesion-relevant visual patterns rather than increased data volume alone. Importantly, these improvements were achieved using a lightweight YOLO-based architecture, highlighting the effectiveness of data-centric and attention-guided strategies in enabling efficient and scalable detection models for clinical deployment. Future work may focus on crossinstitutional validation, extension to pixel-level lesion segmentation, integration of multi-view mammographic consistency and radiologist-in-the-loop evaluation to further improve generalization, interpretability and clinical usability of the proposed framework.

## Declarations


**• Funding**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.


**• Conflict of Interest / Competing Interests**

The authors declare that they have no conflict of interest.

**• Ethics Approval and Consent to Participate**

The study was conducted in accordance with relevant ethical standards. The dataset used in this research is publicly available, and all applicable ethical principles were observed.

**• Consent for Publication**

All necessary consents for publication were obtained.

**• Data Availability**

The original mammographic images used in this study were obtained from the publicly available VinDr-Mammo dataset, hosted on PhysioNet (https://physionet.org/content/vindr-mammo/). The processed and augmented datasets generated during this study, as well as the code developed for data preprocessing, augmentation, and model training, are available from the corresponding author upon reasonable request.

**• Author Contribution**

Zahra Samadi (first author) contributed to problem formulation, theoretical analysis, coding and manuscript writing. Mohammad Norouzi contributed to the theoretical development and manuscript revision. Mona Haddad Zahmatkesh contributed significantly through her expertise in medical concepts, assisted in problem design and participated in manuscript revision. Seyed Omid Shahdi contributed to manuscript revision.

The manuscript has not been submitted to or published in any other journal.

## References


[1] Key, T.J., Verkasalo, P.K., Banks, E.: Epidemiology of breast cancer. The lancet oncology 2(3), 133–140 (2001)
[2] Coughlin, S.S.: Epidemiology of breast cancer in women. Breast Cancer Metastasis and Drug Resistance: Challenges and Progress, 9–29 (2019)
[3] Freihat, O., Sipos, D., Kovacs, A.: Global burden and projections of breast cancer incidence and mortality to 2050: a comprehensive analysis of globocan data. Frontiers in Public Health 13, 1622954 (2025)
[4] Vogel, V.G.: Epidemiology of breast cancer. In: The Breast, pp. 207–218. Elsevier, ??? (2018)
[5] Trimboli, R.M., Giorgi Rossi, P., Battisti, N.M.L., Cozzi, A., Magni, V., Zanardo, M., Sardanelli, F.: Do we still need breast cancer screening in the era of targeted therapies and precision medicine? Insights into Imaging 11(1), 105 (2020)
[6] Huang, S., Westvold, S.J., Soulos, P.R., Fan, J., Winer, E.P., Zhan, H., Lustberg, M.B., Lewin, J., Robinson, T.J., Dinan, M.A.: Screening history, stage at diagnosis, and mortality in screen-detected breast cancer. JAMA Network Open 8(4), 255322–255322 (2025)
[7] De Jesus, C., Moseley, T.W., Diaz, V., Vishwanath, V., Jean, S., Elhatw, A., Pria, H.R.F.D., Chung, H.L., Guirguis, M.S., Patel, M.M.: The benefits of screening mammography. Current Breast Cancer Reports 15(2), 103–107 (2023)
[8] Nguyen, H.T., Nguyen, H.Q., Pham, H.H., Lam, K., Le, L.T., Dao, M., Vu, V.: Vindr-mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. Scientific Data 10(1), 277 (2023)
[9] Shi, J., Li, J., Gao, Y., Chen, W., Zhao, L., Li, N., Tian, J., Li, Z.: The screening value of mammography for breast cancer: an overview of 28 systematic reviews with evidence mapping. Journal of Cancer Research and Clinical Oncology 151(3), 1–20 (2025)
[10] Yang, S., Zhang, C., Zang, Q., Yu, J., Zeng, L., Luo, X., Xing, Y., Pan, X., Li, Q., Liang, X., et al.: Mammo-clustering: A multi-views tri-level information fusion context clustering framework for localization and classification in mammography. arXiv preprint arXiv:2409.14876 (2024)
[11] Mansour, S., Soliman, S., Kansakar, A., Marey, A., Hunold, C., Hanafy, M.M.: Strengths and challenges of the artificial intelligence in the assessment of dense breasts. BJR| Open 4(1), 20220018 (2022)
[12] Huynh, P.T., Jarolimek, A.M., Daye, S.: The false-negative mammogram. Radiographics 18(5), 1137–1154 (1998)
[13] Ekpo, E.U., Alakhras, M., Brennan, P.: Errors in mammography cannot be solved through technology alone. Asian Pacific journal of cancer prevention: APJCP 19(2), 291 (2018)
[14] Zafari, Y., Pan, H., Durak, G., Bagci, U., Rashed, E.A., Mabrok, M.: Mammoclean: Toward reproducible and bias-aware ai in mammography through dataset harmonization. arXiv preprint arXiv:2511.02400 (2025)
[15] Wang, L.: Mammography with deep learning for breast cancer detection. Frontiers in oncology 14, 1281922 (2024)
[16] Radak, M., Lafta, H.Y., Fallahi, H.: Machine learning and deep learning techniques for breast cancer diagnosis and classification: a comprehensive review of medical imaging studies. Journal of Cancer Research and Clinical Oncology 149(12), 10473–10491 (2023)
[17] Gardezi, S.J.S., Elazab, A., Lei, B., Wang, T.: Breast cancer detection and diagnosis using mammographic data: Systematic review. Journal of medical Internet research 21(7), 14464 (2019)
[18] Gao, Y., Lin, J., Zhou, Y., Lin, R.: The application of traditional machine learning and deep learning techniques in mammography: a review. Frontiers in Oncology 13, 1213045 (2023)
[19] Ismailova, A., Abdikerimova, G., Uzakkyzy, N., Muratkhan, R., Aitimov, M., Tergeusizova, A., Beissegul, A.: Development of a feature vector for accurate breast cancer detection in mammographic images. International Journal of Cognitive Computing in Engineering (2025)
[20] Safdar, S., Rizwan, M., Gadekallu, T.R., Javed, A.R., Rahmani, M.K.I., Jawad, K., Bhatia, S.: Bio-imaging-based machine learning algorithm for breast cancer detection. Diagnostics 12(5), 1134 (2022)
[21] Xi, P., Shu, C., Goubran, R.: Abnormality detection in mammography using deep convolutional neural networks. In: 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1–6 (2018). IEEE
[22] Shen, L., Margolies, L., Rothstein, J., McBride, R., Fluder, E., Sieh, W.: Deep learning to improve breast cancer early detection on screening mammography. Clinical Orthopaedics and Related Research (2018)
[23] Prinzi, F., Insalaco, M., Orlando, A., Gaglio, S., Vitabile, S.: A yolo-based model for breast cancer detection in mammograms. Cognitive Computation 16(1), 107–120 (2024)
[24] Palaniappan, D., Jain, R., Premavathi, T., Parmar, K., Ghribi, W., Ahmed,A.M., Ahmad, N.: Yolo in healthcare: A comprehensive review of detectionarchitectures, domain applications, and future innovations. IEEE Access (2025)
[25] Ragab, M.G., Abdulkadir, S.J., Muneer, A., Alqushaibi, A., Sumiea, E.H.,Qureshi, R., Al-Selwi, S.M., Alhussian, H.: A comprehensive systematic review ofyolo for medical object detection (2018 to 2023). IEEE Access 12, 57815–57836(2024)
[26] Mohammed, A.D., Ekmekci, D.: Breast cancer diagnosis using yolo-based multiscaleparallel cnn and flattened threshold swish. Applied Sciences 14(7), 2680(2024)
[27] Kebede, S.R., Waldamichael, F.G., Debelee, T.G., Aleme, M., Bedane, W.,Mezgebu, B., Merga, Z.C.: Dual view deep learning for enhanced breast cancerscreening using mammography. Scientific Reports 14(1), 3839 (2024)
[28] Kolchev, A., Pasynkov, D., Egoshin, I., Kliouchkin, I., Pasynkova, O., Tumakov,D.: Yolov4-based cnn model versus nested contours algorithm in the suspiciouslesion detection on the mammography image: A direct comparison in the realclinical settings. Journal of Imaging 8(4), 88 (2022)
[29] Avcı, H., Karakaya, J.: A novel medical image enhancement algorithm for breastcancer detection on mammography images using machine learning. Diagnostics13(3), 348 (2023)
[30] Kurt, B., Nabiyev, V.V., Turhan, K.: Comparison of enhancement methods formammograms with performance measures. In: e-Health–For Continuity of Care,pp. 486–490. IOS Press, ??? (2014)
[31] Wu, S., Yu, S., Yang, Y., Xie, Y.: Feature and contrast enhancement of mammographicimage based on multiscale analysis and morphology. Computational andmathematical methods in medicine 2013(1), 716948 (2013)
[32] Abdikenov, B., Rakishev, D., Orazayev, Y., Zhaksylyk, T.: Enhancing breastlesion detection in mammograms via transfer learning. Journal of Imaging 11(9),314 (2025)
[33] Islam, W., Jones, M., Faiz, R., Sadeghipour, N., Qiu, Y., Zheng, B.: Improvingperformance of breast lesion classification using a resnet50 model optimized witha novel attention mechanism. Tomography 8(5), 2411–2425 (2022)

[34] Ahmed, H., Nandi, A.K.: Asg-mammonet: An attention-guided framework forstreamlined and interpretable breast cancer classification from mammograms.Frontiers in Signal Processing 5, 1672569 (2025)
[35] Nguyen, H.T., Nguyen, H.Q., Pham, H.H., Lam, K., Le, L.T., Dao, M., Vu, V.:Vindr-mammo: A large-scale benchmark dataset for computer-aided diagnosisin full-field digital mammography. Scientific Data 10(1), 277 (2023) https://doi.org/10.1038/s41597-023-02100-7
[36] Mishra, A.: Contrast Limited Adaptive Histogram Equalization (CLAHE) approach for enhancement of the microstructures of friction stir welded joints. arXiv preprint arXiv:2109.00886. Published August 15, 2021 (2021). https:// arxiv.org/abs/2109.00886
[37] Matsumoto, M.: Noise reduction with complex bilateral filter. The Journal of the Acoustical Society of America 142(6), 3486–3496 (2017) https://doi.org/10. 1121/1.5016807
[38] Xue, J., Cheng, F., Li, Y., Song, Y., Mao, T.: Detection of farmland obstacles based on an improved YOLOv5s algorithm by using CIoU and anchor box scale clustering. Sensors 22(5), 1790 (2022) https://doi.org/10.3390/s22051790
[39] Lin, T.-Y., Goyal, P., Girshick, R.B., He, K., Dollár, P.: Focal Loss for Dense Object Detection (2017). https://arxiv.org/abs/1708.02002
[40] Xu, L., Yan, W., Ji, J.: The research of a novel wog-yolo algorithm for autonomous driving object detection. Scientific reports 13(1), 3699 (2023)
[41] Wu, J.: Anomaly detection in medical via multimodal foundation models. Frontiers in Bioengineering and Biotechnology 13, 1644697 (2025)
[42] Reyes, M., Meier, R., Pereira, S., Silva, C.A., Dahlweid, F.-M., Tengg-Kobligk, H.v., Summers, R.M., Wiest, R.: On the interpretability of artificial intelligence in radiology: challenges and opportunities. Radiology: artificial intelligence 2(3), 190043 (2020)
[43] Hana, T., Tanaka, S., Nejo, T., Takahashi, S., Kitagawa, Y., Koike, T., Nomura, M., Takayanagi, S., Saito, N.: Mining-guided machine learning analyses revealed the latest trends in neuro-oncology. Cancers 11(2), 178 (2019)
[44] Zhu, D.: Pose estimation for health data analysis: advancing ai in neuroscience and psychology. Frontiers in Neurology 16, 1596408 (2025)