



## Research Article

# Intelligent detection of subtle pulmonary abnormalities in CXR images using YOLO-CA-NET framework based on YOLOv11 and coordinate attention

Melika Ghasemi<sup>1</sup> , \* Hamid Reza Ghaffari<sup>1</sup>

1. Department of Computer Engineering, Fe.C., Islamic Azad University, Ferdows, Iran.

Use your device to scan  
and read the article online



**Citation:** Ghasemi M, Ghaffari HR. Intelligent detection of subtle pulmonary abnormalities in CXR images using YOLO-CA-NET framework based on YOLOv11 and coordinate attention. *Journal of Modern Medical Information Science*. 2025; 11(1):201-218. [In Persian]



10.48312/jmis.11.2.904.2

### Article Info:

Received: 2 Jun 2025

Accepted: 20 Sep 2025

Available Online: 20  
Sep 2025

### Key Words:

Thoracic Radiography,  
Subtle Pathologies,  
YOLOv11, Coordinate  
Attention Mechanism,  
Object Detection.

## ABSTRACT

**Introduction:** Chest radiography is a vital tool for disease diagnosis; however, it is susceptible to human error due to the subtlety of findings and their overlap with complex anatomical structures. The objective of this research is to present a new framework, YOLO-CA-NET, to improve the detection accuracy and localization of subtle and small abnormalities in radiographic images using advanced attention mechanisms.

**Methods:** In this study, the YOLOv11 model was upgraded by strategically integrating the Coordinate Attention module into its backbone blocks to extract position-sensitive information and preserve the spatial details of small lesions. For training and evaluation, the standard VinDr-CXR dataset, comprising 15,000 radiographic images with annotations for 14 common types of abnormalities, was utilized. Pre-processing included normalization and the CLAHE technique to enhance the contrast of low-visibility findings.

**Results:** The research findings indicated that the proposed YOLO-CA-NET model achieved a mean Average Precision (mAP) of 0.329 and a mean Recall of 0.514. These results demonstrate a 9.3% increase in precision and a 19% increase in recall compared to the baseline YOLOv11 model. This improvement validates the high potential of the proposed framework in detecting and localizing low-contrast and subtle lesions that might be overlooked by the baseline model.

**Discussion:** Integrating the Coordinate Attention mechanism into the YOLOv11 architecture significantly enhances the model's ability to understand long-range spatial dependencies and precisely localize subtle abnormalities. The proposed framework can serve as an efficient computer-aided diagnosis tool to mitigate human errors arising from missing small lesions in clinical settings.

### \* Corresponding Author:

Dr Hamid Reza Ghaffari

Address: Fe.C., Islamic Azad University, Ferdows, Iran.

E-mail: hamidghaffary53@iau.ac.ir



Copyright © 2025 The Author[s];

This is an open access article distributed under the terms of the Creative Commons Attribution License [CC-BY-NC; <https://creativecommons.org/licenses/by-nc/4.0/legalcode>], which permits use, distribution, and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.



## Extended Abstract

### Introduction:

**C**hest radiography (CXR) is a foundational tool in diagnostic medical imaging, playing a vital role in identifying a wide spectrum of cardiopulmonary diseases. Despite its critical importance, interpreting these images is highly challenging due to the 2D representation of 3D structures, complex organ overlaps, and inherently low soft-tissue contrast. This visual complexity, combined with high workloads and radiologist fatigue, significantly increases the probability of human error in detecting subtle findings. While deep learning has revolutionized medical image analysis, classical architectures primarily focus on image-level classification rather than precise localization. Single-stage object detection models like YOLO offer real-time processing capabilities suitable for clinical environments. However, baseline YOLO models continue to struggle with detecting miniature objects and low-contrast abnormalities. The structural overlap in CXR images causes precise spatial features to weaken in the deep layers of the network. Standard attention mechanisms, such as SE and CAM, attempt to address this but lose precise spatial information during global pooling operations. This deficiency is particularly problematic in medical imaging where the exact location of a subtle lesion is critical. To bridge this gap, this study presents YOLO-CA-NET, a novel framework based on the YOLOv11 architecture integrated with a Coordinate Attention mechanism. The primary objective is to enrich the semantic representation of early network layers to improve the detection accuracy and precise localization of 14 common, subtle pulmonary abnormalities in CXR images..

### Methods:

This applied analytical study utilized the standard VinDr-CXR dataset, comprising 15,000 high-quality radiographic images annotated for 14 clinical abnormalities by expert radiologists. The dataset was split into 80% for training/validation (13,000 images) and 20% for independent testing (3,000 images) using stratified random sampling to maintain class prevalence homogeneity. Preprocessing involved resizing images to 640\*640 pixels, intensity normalization, and applying the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique to enhance

the visibility of obscure tissues. The proposed YOLO-CA-NET model builds upon the YOLOv11 architecture, consisting of a Backbone, Neck (PANet), and Head. The core methodological innovation is the strategic integration of the Coordinate Attention (CA) module into the backbone's bottleneck blocks. Unlike standard models, the CA module factors channel attention into two parallel 1D feature encoding processes along horizontal and vertical axes to embed direction-aware and position-sensitive information. For an input feature tensor  $X \in R^{C \times H \times W}$ , the mechanism aggregates spatial information using 1D global pooling:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq j \leq W} x_c(h, j)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq i \leq H} x_c(i, w)$$

The output is generated by applying the attention weights to the input:

$$y_c(i, j) = x_c(i \times j) \times g_c^h(i) \times g_c^w(j)$$

This placement allows the CA module to act as an active filter, preventing the leakage of information regarding tiny lesions into dense surrounding tissues like ribs and vessels before dimensionality reduction occurs. Implementation was conducted using PyTorch and the Ultralytics library on an NVIDIA GeForce RTX 4090, training for 200 epochs with the AdamW optimizer.

### Results:

The evaluation of the YOLO-CA-NET framework demonstrated superior performance across primary diagnostic metrics compared to both the baseline model and classical architectures. The proposed model achieved an overall mean Average Precision (mAP@0.5) of 0.329 and a mean recall of 0.514 on the independent test set. As detailed in the comparative analysis, this represents a 9.3% increase in precision and a 19% increase in recall compared to the baseline YOLOv11. An ablation study (Table 4) analyzing attention mechanisms confirmed that integrating the CA module (mAP 0.329, Precision 0.341) significantly outperformed configurations using standard SE blocks (mAP 0.319) or CBAM



(mAP 0.314), proving its unique efficiency in preserving the spatial coordinates of low-contrast lesions. Furthermore, computational cost analysis (Table 5) revealed that the architectural complexity remains highly optimized. YOLO-CA-NET requires only 9.6M parameters and 22.1 GFLOPs, representing a negligible overhead of 0.2M parameters compared to the baseline, while reducing parameters by over 76% and GFLOPs by 83% compared to Faster R-CNN (mAP 0.241). A paired bootstrap test confirmed the statistical significance of this improvement ( $P < 0.05$ ). Qualitatively, visual evaluation (Figure 4) demonstrated the model's high effectiveness in real-world scenarios. The generated bounding boxes precisely localized both large structural abnormalities, such as cardiomegaly, and subtle, low-contrast diffuse nodules simultaneously within a highly complex anatomical background, without requiring heavy auxiliary preprocessing.

## Conclusion:

This study successfully developed and evaluated a highly efficient artificial intelligence model for the automated detection of 14 common thoracic abnormalities. The YOLO-CA-NET framework achieved an optimal balance between diagnostic accuracy and computational cost, making it an ideal candidate for real-world clinical deployment, such as integration into hospital PACS systems as a primary screening tool. The significant improvement in the recall rate confirms that embedding precise positional information during feature learning is critical for identifying subtle pulmonary pathologies and reducing human error caused by radiologist fatigue. Despite its achievements, the model encountered challenges with extremely subtle and heavily imbalanced classes, such as atelectasis (0.71%) and pneumothorax (0.64%), where miniature spatial features decay during successive convolutional downsampling operations. Furthermore, relying on a single dataset highlights the need for broader demographic validation. Future research should prioritize external validation using multi-center datasets like MIMIC-CXR to ensure cross-border generalizability. Addressing class imbalance through specialized loss functions, such as balanced Focal Loss, and implementing explainable AI methods like Grad-CAM for generating heatmap decision visualizations are key next steps. Ultimately, merging the localized feature extraction capabilities of YOLO with the long-range depen-

dency modeling of Vision Transformers presents a highly promising avenue for further enriching spatial features and capturing miniature pathologies.



## مقاله پژوهشی

# تشخیص هوشمند ناهنجاری‌های ظریف ریوی در تصاویر رادیوگرافی قفسه سینه با استفاده از چارچوب YOLO-CA-NET مبتنی بر YOLOv11 و مکانیزم توجه هماهنگ

ملیکا قاسمی<sup>۱</sup>، \* حمیدرضا غفاری<sup>۱</sup> <sup>ID</sup>

۱. گروه مهندسی کامپیوتر، واحد فردوس، دانشگاه آزاد اسلامی، فردوس، ایران.



**Citation:** Ghasemi M, Ghaffari HR. Intelligent detection of subtle pulmonary abnormalities in CXR images using YOLO-CA-NET framework based on YOLOv11 and coordinate attention. *Journal of Modern Medical Information Science*. 2025; 11(1):201-218. [In Persian]

doi 10.48312/jmis.11.2.904.2

## چکیده

**هدف:** رادیوگرافی قفسه سینه ابزاری حیاتی برای تشخیص بیماری‌ها است، اما به دلیل ظرافت یافته‌ها و هم‌پوشانی با ساختارهای آناتومیک پیچیده، مستعد خطای انسانی می‌باشد. هدف این پژوهش، ارائه چارچوب جدیدی به نام YOLO-CA-NET جهت بهبود دقت تشخیص و محلی‌سازی ناهنجاری‌های ظریف و کوچک در تصاویر رادیوگرافی با استفاده از مکانیزم‌های توجه پیشرفته است.

**روش‌ها:** در این مطالعه، مدل YOLOv11 با ادغام استراتژیک ماژول توجه هماهنگ در بلوک‌های ستون فقرات ارتقا یافت تا اطلاعات حساس به موقعیت جهت حفظ جزئیات فضایی ضایعات ریز استخراج شود. برای آموزش و ارزیابی، از مجموعه داده استاندارد VinDr-CXR شامل ۱۵۰۰۰ تصویر رادیوگرافی با حاشیه‌نویسی ۱۴ نوع ناهنجاری استفاده شد. عملیات پیش‌پردازش شامل نرمال‌سازی و تکنیک CLAHE جهت بهبود کنتراست یافته‌های محو اعمال گردید.

**نتایج:** یافته‌های پژوهش نشان داد که مدل پیشنهادی YOLO-CA-NET به میانگین دقت متوسط (mAP) ۰/۳۲۹ و میانگین بازیابی ۰/۵۱۴ دست یافته است. این نتایج در مقایسه با مدل پایه YOLOv11، افزایش ۹/۳ درصدی در دقت و ۱۹ درصدی در بازیابی را نشان می‌دهد؛ این بهبود، پتانسیل بالای چارچوب پیشنهادی را در کشف و محلی‌سازی ضایعات کم‌کنتراست و ظریف که ممکن است در مدل پایه نادیده گرفته شوند، اثبات می‌کند.

**نتیجه‌گیری:** ادغام مکانیزم توجه هماهنگ در معماری YOLOv11، توانایی مدل را در درک وابستگی‌های فضایی دوربرد و محلی‌سازی دقیق ناهنجاری‌های ظریف به‌طور چشمگیری ارتقا می‌دهد. چارچوب پیشنهادی می‌تواند به عنوان ابزاری کمک‌تشخیصی کارآمد برای کاهش خطای انسانی ناشی از نادیده گرفتن ضایعات کوچک در محیط‌های بالینی مورد استفاده قرار گیرد.

## اطلاعات مقاله :

تاریخ دریافت: ۱۳ فرورداد ۱۴۰۴

تاریخ پذیرش: ۳۰ شهریور ۱۴۰۴

تاریخ انتشار: ۳۰ شهریور ۱۴۰۴

## کلیدواژه‌ها :

رادیوگرافی قفسه سینه، ناهنجاری‌های ظریف، YOLOv11، مکانیزم توجه هماهنگ، تشخیص اشیاء.

## \*نویسنده مسئول:

دکتر حمیدرضا غفاری

نشانی: واحد فردوس، دانشگاه آزاد اسلامی، فردوس، ایران.

پست الکترونیک: hamidghaffary53@iau.ac.ir



Copyright © 2025 The Author[s].

This is an open access article distributed under the terms of the Creative Commons Attribution License [CC-BY-NC: <https://creativecommons.org/licenses/by-nc/4.0/legal-code.en>], which permits use, distribution, and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

## مقدمه:

تصاویر رادیوگرافی قفسه سینه<sup>۱</sup> به عنوان پایه و اساس تصویربرداری تشخیصی پزشکی، نقشی حیاتی در شناسایی طیف وسیعی از بیماری‌های قلبی-ریوی ایفا می‌کنند. بر اساس گزارش رسمی سازمان جهانی بهداشت، سالانه به طور تخمینی بیش از ۶/۳ میلیارد تصویربرداری تشخیصی پزشکی در سراسر جهان انجام می‌شود [۱]. در این میان، بر اساس مطالعات جامع اپیدمیولوژیک منتشرشده در مجله انجمن پزشکی آمریکا، رادیوگرافی (اشعه ایکس) کماکان به عنوان پایه و اساس تصویربرداری تشخیصی، حدود ۵۵ درصد از کل این حجم عملیاتی را به خود اختصاص می‌دهد [۲]. این تصاویر، اطلاعات حیاتی درباره اندازه، شکل و وضعیت اندام‌های داخلی از جمله ریه‌ها، قلب، دنده‌ها و عروق خونی فراهم می‌سازد [۳]. با این حال، تفسیر این تصاویر به دلیل نمایش ساختارهای سه‌بعدی روی صفحه دوبعدی، همپوشانی ارگان‌ها و کنتراست پایین بافت نرم، بسیار چالش برانگیز است. این پیچیدگی بصری در کنار حجم کاری بالا و خستگی رادیولوژیست‌ها، احتمال خطای انسانی در تشخیص یافته‌های ظریف را افزایش می‌دهد [۴]. در نتیجه، نیاز مبرمی به توسعه سیستم‌های کمک تشخیصی خودکار وجود دارد تا دقت و کارایی فرآیندهای رادیولوژیک ارتقا یابد.

در دهه اخیر، یادگیری عمیق تحولی شگرف در تحلیل تصاویر پزشکی ایجاد کرده است [۵]. معماری‌های کلاسیک مانند VGG، ResNet و DenseNet اگرچه در استخراج ویژگی‌های قدرتمند موفق بوده‌اند، اما عمدتاً بر طبقه‌بندی در سطح تصویر تمرکز دارند و قادر به محلی‌سازی دقیق ناحیه آسیب‌شناسی نیستند [۶-۱۰]. در مقابل، الگوریتم‌های تشخیص اشیاء با تعیین مکان دقیق ناهنجاری از طریق کادرهای مرزی، ارزش بالینی بیشتری برای برنامه‌ریزی درمان دارند. الگوریتم‌های دو مرحله‌ای تشخیص اشیاء مانند Faster R-CNN به دلیل دقت بالا محبوبیت دارند، اما سرعت استنتاج پایین‌تری نسبت به معماری‌های تک‌مرحله‌ای مانند YOLO [۱۱، ۱۲] دارند. در

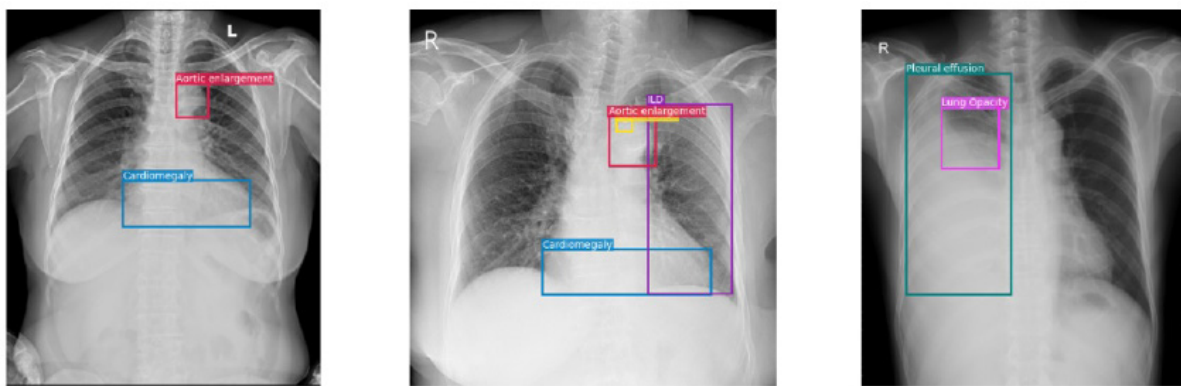
این میان، خانواده الگوریتم‌های YOLO به دلیل معماری یک‌مرحله‌ای و سرعت پردازش بلادرنگ، به عنوان یک راه‌حل پیشرفته برای محیط‌های بالینی شناخته می‌شوند [۱۳]. با این حال، مدل‌های پایه YOLO همچنان در تشخیص اشیاء بسیار کوچک و ناهنجاری‌های با کنتراست پایین که ماهیت ظریفی دارند، با چالش مواجه هستند. همان‌طور که در شکل ۱ نشان داده شده است، همپوشانی ساختاری در تصاویر CXR باعث می‌شود ویژگی‌های مکانی دقیق در لایه‌های عمیق شبکه تضعیف شوند [۱۴].

برای غلبه بر این محدودیت‌ها و عبور از بازنمایی‌های ثابت، پژوهشگران به مکانیزم‌های توجه روی آورده‌اند که به شبکه اجازه می‌دهد به طور تطبیقی بر مناطق مهم تصویر تمرکز کند [۱۵]. با این حال، مکانیزم‌های اولیه مانند SE و CBAM، اگرچه کارآمد هستند، اما اطلاعات مکانی دقیق را در فرآیند Pooling سراسری از دست می‌دهند [۱۶، ۱۷]. این نقیصه در تصاویر پزشکی که در آن موقعیت دقیق یک ناهنجاری ظریف اهمیت حیاتی دارد، دوچندان می‌شود. مکانیزم توجه هماهنگ که توسط هوو و همکاران معرفی شد، این محدودیت را با تجزیه توجه کانالی به دو فرآیند کدگذاری یک‌بعدی در امتداد محورهای افقی و عمودی برطرف می‌کند [۱۸]. این رویکرد به مدل اجازه می‌دهد تا وابستگی‌های فضایی دوربرد را ثبت کرده و هم‌زمان موقعیت دقیق اشیاء را حفظ کند؛ ویژگی متمایزی که برای محلی‌سازی دقیق ناهنجاری‌های ظریف ریوی ضروری است.

همچنین در سال‌های اخیر، تلاش‌های متعددی برای تلفیق مدل‌های خانواده YOLO با مکانیزم‌های توجه جهت تحلیل تصاویر پزشکی صورت گرفته است. استفاده از ماژول‌های مبتنی بر CBAM و ساختارهای مبتنی بر Transformer توانسته‌اند ویژگی‌های کانالی را بهبود بخشند؛ با این حال، اکثر این رویکردها به دلیل استفاده از عملیات Pooling سراسری، در حفظ اطلاعات فضایی و جهت‌دار ضایعات ظریف و ریز ریوی (نظیر نودول‌ها) با چالش مواجه هستند [۱۶، ۱۹-۲۲]. این نقیصه و ناتوانی در تفکیک دقیق مرز آسیب‌شناسی‌های کم‌کنتراست از

1 Chest X-Ray: CXR

با توجه به ضرورت ارتقای دقت در تحلیل تصاویر پزشکی و غلبه بر چالش‌های تشخیص ضایعات کم‌کنتراست، هدف اصلی این پژوهش تشخیص هوشمند ناهنجاری‌های ظریف ریوی در تصاویر رادیوگرافی قفسه سینه با استفاده از چارچوب YOLO-CA-NET مبتنی بر YOLOv11 و مکانیزم توجه هماهنگ است [۱۸، ۲۳]. در این راستا، کارایی چارچوب پیشنهادی جهت شناسایی و محلی‌سازی دقیق ۱۴ ناهنجاری شایع ریوی بر روی مجموعه داده استاندارد VinDr-CXR مورد سنجش، ارزیابی و تحلیل قرار گرفته است [۱۴].



شکل ۱: نمونه تصاویر رادیوگرافی قفسه سینه و برچسب‌گذاری ناهنجاری‌ها از مجموعه داده مرجع VinDr-CXR

پژوهش حاضر، به دلیل تمرکز ویژه بر وظیفه تشخیص شیء و محلی‌سازی دقیق ناهنجاری‌های ظریف ریوی در لایه اشیاء صورت گرفته است. سایر مجموعه‌های داده معروف و بزرگ تصویربرداری قفسه سینه مانند CheXpert یا ChestX-ray14 (NIH) عمدتاً برای وظایف طبقه‌بندی در سطح تصویر توسعه یافته‌اند؛ به طوری که دیتاست CheXpert کاملاً فاقد کادرهای مرزی فضایی است و دیتاست ChestX-ray14 نیز تنها برای زیرمجموعه بسیار کوچکی شامل ۸۸۰ تصویر (کمتر از ۱ درصد کل داده‌ها) مختصات مکان‌مند ارائه داده است [۲۴، ۲۵]. در مقابل، این فضا در مجموعه داده VinDr-CXR با ارائه کادرهای مرزی جامع، دقیق و تاییدشده توسط چندین رادیولوژیست متبحر برای تمامی نمونه‌های پاتولوژیک برطرف شده و بستری بهینه را برای توسعه مدل پیشنهادی فراهم ساخته است.

در مرحله تقسیم داده‌ها، جهت تضمین ارزیابی بدون سوگیری، کل مجموعه داده نهایی با نسبت استاندارد ۸۰ به

بافت‌های آناتومیک هم‌پوشان، گپ پژوهشی بزرگی را در سیستم‌های کمک تشخیصی خودکار ایجاد کرده است.

نوآوری محوری این پژوهش، غنی‌سازی معنایی لایه‌های اولیه شبکه از طریق ادغام استراتژیک توجه هماهنگ است [۱۸]. این راهبرد با تجزیه توجه کانالی به دو فرآیند کدگذاری ویژگی یک‌بعدی موازی، مدل را قادر می‌سازد تا برخلاف مدل‌های کلاسیک، وابستگی‌های فضایی دوربرد را ثبت کرده و مکان دقیق ناهنجاری‌های بسیار ریز را در سراسر شبکه حفظ و تقویت کند.

## مواد و روش‌ها:

این پژوهش یک مطالعه تحلیلی-کاربردی در حوزه هوش مصنوعی پزشکی است که با بهره‌گیری از داده‌های ثانویه انجام شده است. جامعه مورد مطالعه شامل تصاویر رادیوگرافی دیجیتال قفسه سینه است که از مجموعه داده استاندارد VinDr-CXR (منتشر شده توسط انستیتو بیگ‌دیتای Vingroup) استخراج گردید [۱۴]. از میان ۱۸/۰۰۰ تصویر موجود در این بانک اطلاعاتی اولیه، تعداد ۳/۰۰۰ تصویر به دلیل کیفیت نامناسب تصویربرداری، نویز شدید ساختاری یا نقص در حاشیه‌نویسی رادیولوژیست‌ها از چرخه تحلیل حذف گردید و در نهایت، زیرمجموعه‌ای شامل ۱۵/۰۰۰ تصویر رادیوگرافی با کیفیت بالا که ۱۴ ناهنجاری شایع بالینی را پوشش می‌داد، به عنوان نمونه نهایی انتخاب شد.

انتخاب این مجموعه داده به عنوان بستر استاندارد

این رویکرد کلیدی تضمین می‌کند که توزیع فراوانی و نسبت شیوع هر یک از ۱۴ کلاس ناهنجاری ریوی (به‌ویژه پاتولوژی‌های ظریف و کم‌تکرار) در هر دو مجموعه آموزش و تست کاملاً همگن، متوازن و مشابه یکدیگر حفظ شده است. جزئیات این توزیع آماری در جدول ۱ ارائه گردیده است.

۲۰ درصد تفکیک گردید؛ به طوری که ۸۰ درصد از داده‌ها (معادل ۱۲/۰۰۰ تصویر) برای فرآیند آموزش و اعتبارسنجی درونی و ۲۰ درصد مابقی (معادل ۳/۰۰۰ تصویر) به عنوان مجموعه آزمون مستقل جهت ارزیابی نهایی کارایی مدل اختصاص یافت. فرآیند تقسیم تصاویر با استفاده از راهبرد نمونه‌گیری تصادفی طبقه‌بندی‌شده پیاده‌سازی شد.

| جدول ۱: توزیع آماری یافته‌های پاتولوژیک در مجموعه داده‌های آموزش و آزمون (مقادیر برحسب درصد شیوع گزارش شده‌اند) |                    |                               |                  |
|-----------------------------------------------------------------------------------------------------------------|--------------------|-------------------------------|------------------|
| کلاس ناهنجاری                                                                                                   | برچسب اصلی         | مجموعه آموزش / اعتبارسنجی (%) | مجموعه آزمون (%) |
| بزرگی آنورت                                                                                                     | Aortic Enlargement | ٪ ۱۹/۹                        | ٪ ۱۹/۹۹          |
| آتلکتازی                                                                                                        | Atelectasis        | ٪ ۰/۷۱                        | ٪ ۰/۹۵           |
| کلسیفیکاسیون                                                                                                    | Calcification      | ٪ ۲/۷۱                        | ٪ ۲/۱۶           |
| کاردیومگالی (بزرگ‌شدگی قلب)                                                                                     | Cardiomegaly       | ٪ ۱۵/۰۸                       | ٪ ۱۴/۰۷          |
| کانسالیدیشن (تراکم ریوی)                                                                                        | Consolidation      | ٪ ۱/۶۹                        | ٪ ۱/۰۴           |
| افیوژن پلورال (تجمع مایع)                                                                                       | Pleural Effusion   | ٪ ۶/۷                         | ٪ ۶/۷۸           |
| فیبروز ریوی                                                                                                     | Pulmonary Fibrosis | ٪ ۱۲/۸                        | ٪ ۱۳/۶           |
| بیماری ریوی بین‌بافتی                                                                                           | ILD                | ٪ ۲/۶۶                        | ٪ ۲/۶۶           |
| ارتشاح ریوی (اینفیلتراسیون)                                                                                     | Infiltration       | ٪ ۳/۵۴                        | ٪ ۳              |
| کدورت ریوی (اوپاسیتی)                                                                                           | Lung Opacity       | ٪ ۶/۸۴                        | ٪ ۷/۲۳           |
| گره / توده ریوی                                                                                                 | Nodule/ Mass       | ٪ ۷/۱۴                        | ٪ ۸/۶۱           |
| سایر ناهنجاریها                                                                                                 | Other Lesion       | ٪ ۶/۱                         | ٪ ۶/۱۷           |
| ضخیم‌شدگی پلورال                                                                                                | Pleural Thickening | ٪ ۱۳/۵۱                       | ٪ ۱۳/۲۳          |
| پنوموتوراکس (تجمع هوا)                                                                                          | Pneumothorax       | ٪ ۰/۶۴                        | ٪ ۰/۵            |

سرور محاسباتی مجهز به پردازنده گرافیکی NVIDIA GeForce RTX 4090 با ۲۴ گیگابایت حافظه اختصاصی و ۳۲ گیگابایت رم، تحت چارچوب PyTorch و کتابخانه رسمی Ultralytics به اجرا درآمد.

ابزار اصلی گردآوری و تحلیل داده‌ها در این تحقیق، مدل طراحی شده YOLO-CA-NET است که بر پایه معماری نوین YOLOv11 توسعه یافته است. همانطور که در شکل ۲ نشان داده شده، این معماری از سه جزء اصلی تشکیل شده است [۲۳، ۲۸]:

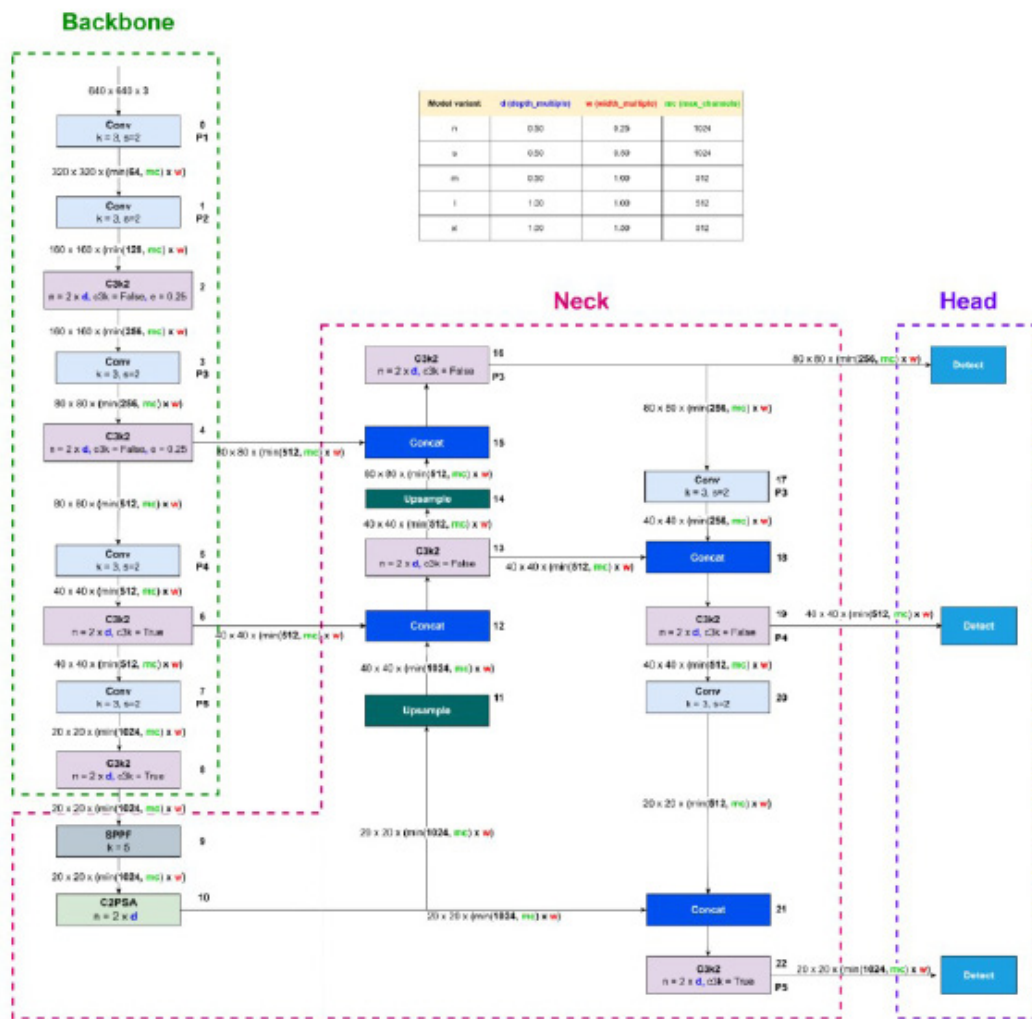
\* ستون فقرات: یک شبکه به سبک جزئی بین مرحله‌ای که از بلوک‌های C3K2 و C2PSA برای استخراج ویژگی کارآمد استفاده می‌کند.

شبیه‌سازی و پیاده‌سازی معماری پیشنهادی در این پژوهش با استفاده از زبان برنامه‌نویسی پایتون (نسخه ۳.۱۰) و فریم‌ورک محبوب و پیشرفته یادگیری عمیق PyTorch انجام شده است [۲۶]. جهت آموزش، تنظیم و ارزیابی مدل، از کتابخانه رسمی، استاندارد و متن‌باز شرکت Ultralytics (نسخه ۱۱/۰) برای معماری YOLOv11 استفاده گردید [۲۷]. بهره‌گیری از پیاده‌سازی رسمی و استاندارد توسعه‌یافته توسط شرکت مالک اثر (Ultralytics)، علاوه بر تضمین صحت کامل عملکرد الگوریتم‌ها، توابع زیان و لایه‌های شبکه، پایداری فرآیند آموزش را حفظ کرده و قابلیت بازتولید دقیق نتایج را برای پژوهش‌های آتی تضمین می‌کند.

تمامی مراحل پیاده‌سازی و پردازش داده‌ها بر روی یک

\* سر: این بخش نهایی، نقشه‌های ویژگی پالایش شده از Neck را دریافت کرده و به طور مستقل کادرهای مرزی، احتمالات کلاس و امتیازات وجود شیء را پیش‌بینی می‌کند.

\* گردن: یک شبکه تجمیع مسیر (PANet) که ویژگی‌ها را از لایه‌های مختلف به صورت بالا-پایین و پایین-بالا ترکیب کرده و با استفاده از ماژول SPFF، یک هرم ویژگی چندمقیاسی غنی ایجاد می‌کند.



شکل ۲: معماری YOLOv11

را حفظ نماید.

با توجه به یک تانسور ویژگی  $X \in R^{C \times H \times W}$  بند CA شامل دو مرحله اصلی است [۱۸]:

۱. تعبیه اطلاعات هماهنگ: ماژول ابتدا ویژگی‌ها را در امتداد جهت‌های افقی و عمودی با استفاده از دو عملیات ادغام سراسری یک‌بعدی تجمیع می‌کند:

این عملیات دو بردار ویژگی تولید می‌کنند که اطلاعات

نوآوری محوری این مطالعه در ادغام استراتژیک ماژول توجه هماهنگ در لایه‌های اولیه ستون فقرات شبکه نهفته است. این ماژول با تجزیه توجه کانالی به دو فرآیند کدگذاری ویژگی یک‌بعدی مطابق معادلات ۱ تا ۴ در امتداد محورهای افقی و عمودی، اطلاعات آگاه از جهت و حساس به موقعیت را در نقشه‌های ویژگی تعبیه می‌کند. این امر به مدل اجازه می‌دهد تا برخلاف مدل‌های کلاسیک که در فرآیند کاهش ابعاد جزئیات ریز را از دست می‌دهند، مکان دقیق ناهنجاری‌های ظریف و وابستگی‌های فضایی دوربرد

این جایگذاری هدفمند تضمین می‌کند که مدل از همان مراحل اولیه استخراج ویژگی، نمایش‌های آگاه از فضا را در مقیاس‌های متعدد یاد می‌گیرد. با غنی‌سازی ویژگی‌ها در عمق ستون فقرات، اطلاعاتی که به اجزای گردن و سر منتقل می‌شود، از نظر مکانی دقیق‌تر و از نظر معنایی متمایزتر است. این رویکرد به ویژه برای وظیفه تشخیص ناهنجاری در تصاویر CXR که در آن آسیب‌شناسی‌ها می‌توانند کوچک، با کنتراست پایین یا همپوشان با ساختارهای آناتومیک دیگر باشند، بسیار مؤثر است. در واقع، با این کار، مدل یاد می‌گیرد که به طور انتخابی بر روی مناطق مرتبط تمرکز کرده و ویژگی‌های مکانی حیاتی را در سراسر شبکه حفظ و تقویت کند.

جهت اطمینان از روایی ابزار، از برچسب‌گذاری‌های مرجع که توسط تیمی از رادیولوژیست‌های مجرب با استفاده از کادرهای محدودکننده دقیق انجام شده بود، بهره گرفته شد. برای ارتقای پایایی تشخیص در شرایط نوری متفاوت، تصاویر ورودی به ابعاد  $640 \times 640$  پیکسل تغییر یافته و مقادیر شدت روشنایی آن‌ها نرمال‌سازی شد؛ همچنین تکنیک برابری هیستوگرام تطبیقی با محدودیت کنتراست جهت برجسته‌سازی جزئیات بافت‌های محو اعمال گردید. مدل پیشنهادی با استفاده از وزن‌های پیش‌آموزش‌دیده COCO و اعمال استراتژی داده‌افزایی پویا شامل موزاییک و چرخش افقی، طی ۲۰۰ دوره با بهینه‌ساز AdamW تنظیم دقیق شد.

در نهایت، تجزیه و تحلیل آماری نتایج با استفاده از معیارهای استاندارد میانگین دقت متوسط در آستانه  $0.5 = \text{IoU}$ ،  $(\text{mAP}@0.5)$  و میانگین بازیابی در نرخ‌های ثابت مثبت کاذب انجام گرفت. جهت تعیین معناداری آماری تفاوت عملکرد بین مدل پیشنهادی و مدل پایه، از آزمون بوت‌استرپ زوجی با نمونه‌برداری مجدد و بازه اطمینان ۹۵ درصد استفاده شد. در تمامی تحلیل‌ها، سطح معنادار  $P < 0.05$  در نظر گرفته شد تا اطمینان حاصل شود که بهبودهای به‌دست‌آمده در قدرت تشخیصی مدل، از نظر آماری مستدل و غیرتصادفی هستند.

موقعیتی را در امتداد دو محور فضایی کدگذاری می‌کنند.

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq j \leq W} x_c(h, j) \quad (1)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq i \leq H} x_c(i, w) \quad (2)$$

۲. تولید توجه هماهنگ: دو بردار ویژگی به هم متصل شده و از یک لایه کانولوشنی  $1 \times 1$  مشترک  $F_1$  و یک تابع فعال‌سازی غیرخطی ( $\delta$ ) عبور می‌کنند تا یک نقشه ویژگی میانی  $f$  تولید شود. سپس  $f$  به دو تانسور جداگانه تقسیم شده و توسط دو لایه کانولوشنی  $1 \times 1$  دیگر ( $F_w, F_h$ ) و یک تابع سیگموئید ( $\sigma$ ) تبدیل می‌شوند تا وزن‌های توجه نهایی تولید شوند:

$$g^w = \sigma(F_w(f^w)), \quad g^h = \sigma(F_h(f^h)) \quad (3)$$

خروجی نهایی  $Y$  با اعمال این وزن‌های توجه به تانسور ورودی به دست می‌آید:

$$y_c(i, j) = x_c(i \times j) \times g_c^h(i) \times g_c^w(j) \quad (4)$$

به‌منظور بهره‌برداری حداکثری از اطلاعات مکانی، ماژول توجه هماهنگ به‌صورت استراتژیک در انتهای بلوک‌های Bottleneck در ستون فقرات شبکه جای‌گذاری شد. انتخاب این نقطه به دو دلیل کلیدی صورت گرفت: اولاً، پس از استخراج ویژگی‌های سطح بالا در لایه‌های عمیق‌تر، نقشه ویژگی‌ها دارای غنای معنایی کافی برای تمرکز بر نواحی کلیدی هستند، اما در عین حال، وضوح فضایی هنوز به‌طور کامل در اثر کانولوشن‌های متوالی از دست نرفته است. ثانیاً، قرارگیری ماژول CA قبل از ورود به لایه‌های Neck (جایی که ویژگی‌ها برای پیش‌بینی نهایی تجمع می‌شوند)، به شبکه اجازه می‌دهد تا پیش از کاهش ابعاد، وزن‌های توجه را بر روی مناطق حساس (مانند لبه‌های ظریف ضایعات) اعمال کند. این استقرار موجب شد تا ماژول CA به عنوان یک فیلتر فعال عمل کرده و از نشت اطلاعات مربوط به ضایعات ریز در بافت‌های متراکم (مانند عروق و دنده‌ها) جلوگیری نماید.

سخت‌افزاری و نرم‌افزاری به صورت یکپارچه در جدول ۲ ارائه شده است. تمامی این تنظیمات در طول فرآیند آموزش و ارزیابی بر روی ۱۵/۰۰۰ تصویر رادیوگرافی قفسه سینه به طور ثابت حفظ گردیده‌اند.

جهت تضمین شفافیت علمی و فراهم نمودن امکان بازتولید دقیق نتایج این پژوهش برای تحقیقات آتی، تمامی جزئیات مرتبط با تنظیمات پیش‌پردازش تصاویر، هایپرپارامترهای کلیدی آموزش مدل پیشنهادی YOLO-CA-NET، استراتژی‌های داده‌افزایی و مشخصات زیرساخت

| جدول ۲: مشخصات جامع هایپرپارامترها، تنظیمات پیش‌پردازش و محیط پیاده‌سازی مدل پیشنهادی |                                                                                                               |                                                                             |
|---------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| دسته‌بندی فنی                                                                         | پارامتر / مشخصه متدولوژیک                                                                                     | مقدار / تنظیمات اعمال شده                                                   |
| پیش‌پردازش داده‌ها                                                                    | ابعاد تصویر ورودی<br>تکنیک بهبود کنتراست موضعی<br>نرمال‌سازی شدت روشنایی                                      | ۶۴۰*۶۴۰ پیکسل<br>CLAHE<br>مقیاس‌دهی خطی در بازه [۱/۰]                       |
| هایپرپارامترهای مدل                                                                   | تعداد دوره‌های آموزش<br>اندازه دسته<br>الگوریتم بهینه‌ساز<br>نرخ یادگیری اولیه<br>ضریب تضعیف وزن<br>تابع زیان | ۲۰۰ دوره<br>۱۶<br>AdamW<br>۰/۰۰۱<br>۰/۰۰۰۵<br>CIoU Loss + BCE Loss          |
| داده‌افزایی پویا                                                                      | استراتژی ترکیب تصاویر<br>تغییرات هندسی پویا                                                                   | موزاییک<br>چرخش افقی، تغییر مقیاس                                           |
| محیط نرم‌افزاری                                                                       | فریم‌ورک اصلی یادگیری عمیق<br>کتابخانه توسعه معماری                                                           | PyTorch (نسخه ۲/۱) / زبان Python ۳/۱۰<br>Ultralytics YOLOv11<br>(نسخه رسمی) |
| زیرساخت سخت‌افزاری                                                                    | پردازنده گرافیکی<br>حافظه موقت سیستم (RAM)                                                                    | NVIDIA GeForce RTX 4090 (24 GB VRAM)<br>۳۲ گیگابایت DDR۵                    |

### یافته‌ها:

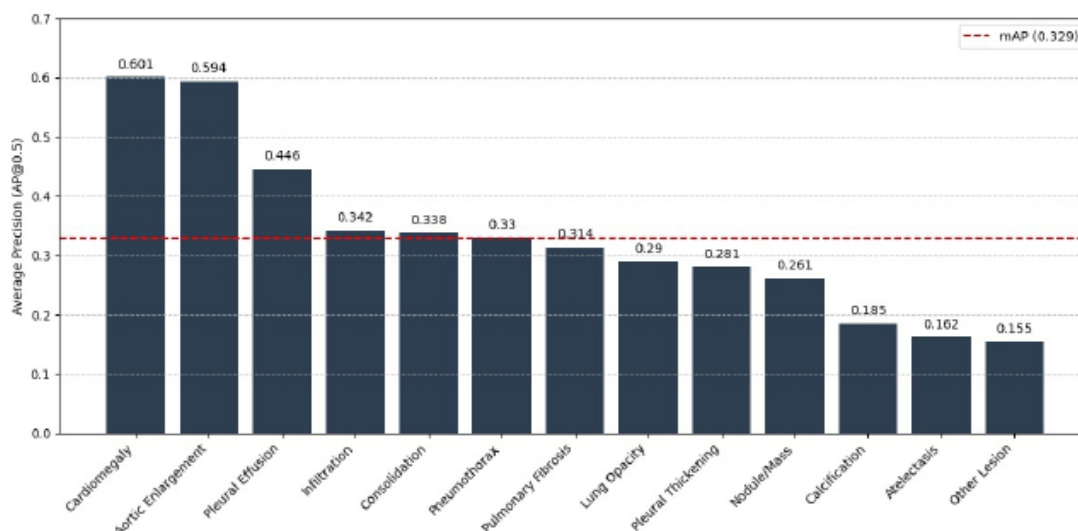
کلاس‌بندی و مطالعه حذفی ارائه می‌گردد. یافته‌های اولیه نشان می‌دهد که مدل ارتقایافته در تمامی سنج‌های تشخیصی نسبت به مدل پایه و سایر معماری‌های کلاسیک برتری دارد.

در این بخش، نتایج حاصل از ارزیابی مدل پیشنهادی YOLO-CA-NET در سه سطح عملکرد کلی، تحلیل

| جدول ۳: مقایسه عملکرد کلی بر روی مجموعه آزمون VinDr-CXR |         |                 |                |                |                 |
|---------------------------------------------------------|---------|-----------------|----------------|----------------|-----------------|
| مدل                                                     | mAP@0.5 | بازیابی FP@0.05 | بازیابی FP@0.1 | بازیابی FP@0.2 | میانگین بازیابی |
| Faster R-CNN (ResNet50-FPN)                             | ۰/۲۴۱   | ۰/۴۲۵           | ۰/۴۲۸          | ۰/۴۵۷          | ۰/۴۳۷           |
| Sparse R-CNN                                            | ۰/۲۴۹   | ۰/۴۲۹           | ۰/۴۴۱          | ۰/۴۵۲          | ۰/۴۴۱           |
| YOLOv8s                                                 | ۰/۲۶۵   | ۰/۴۳۳           | ۰/۴۴۷          | ۰/۴۵۹          | ۰/۴۴۷           |
| YOLOv11 پایه                                            | ۰/۳۰۱   | ۰/۴۴۲           | ۰/۴۶۸          | ۰/۴۷۲          | ۰/۴۶۱           |
| YOLO-CA-NET (پیشنهادی)                                  | ۰/۳۲۹   | ۰/۴۶۱           | ۰/۵۱۷          | ۰/۵۶۲          | ۰/۵۱۴           |

پایه نشان می‌دهد. این جهش در نرخ بازیابی نشان‌دهنده توانایی بالای مدل در کشف ضایعات کم‌کنتراست و ظرفیتی است که در مدل پایه نادیده گرفته می‌شدند. جهت درک بهتر و ملموس‌تر توزیع سنج‌های عملکردی در کلاس‌های مختلف پاتولوژی، میزان دقت تفکیکی مدل پیشنهادی در شکل ۳ ترسیم شده است.

تحلیل مقایسه‌ای ارائه شده در جدول ۳ حاکی از آن است که مدل YOLO-CA-NET به بالاترین میزان دقت متوسط معادل ۰/۳۲۹ دست یافته است. این عملکرد در سنج‌های بازیابی نیز تکرار شده و در نرخ مثبت کاذب ۰/۲، مدل موفق به شناسایی ۵۶/۲ درصد از ناهنجاری‌ها شده است که بهبودی معادل ۱۹ درصد را نسبت به مدل



شکل ۳: مقایسه میانگین دقت متوسط (mAP@0.5) در ناهنجاری‌های مختلف ریوی بر روی مجموعه داده VinDr-CXR

مؤلفه‌های معماری پیشنهادی و توجیه فنی ادغام ماژول توجه هماهنگ (CA) در ستون فقرات شبکه YOLOv11، یک مطالعه حذفی جامع طراحی و اجرا شد. در این ارزیابی، عملکرد مدل پایه YOLOv11 با حالت‌هایی که مکانیزم‌های توجه استاندارد دیگر نظیر SE و CBAM به آن افزوده شده‌اند، مورد مقایسه قرار گرفت تا مشخص شود آیا بهبود کارایی صرفاً ناشی از افزودن یک ماژول توجه دلخواه بوده یا به ویژگی‌های ساختاری لایه پیشنهادی وابسته است [۱۶، ۲۹]. نتایج کمی حاصل از این ارزیابی بر روی مجموعه داده آزمون مستقل VinDr-CXR در جدول ۴ گزارش شده است.

بر اساس یافته‌های مندرج در جدول ۴، مشخص است که الحاق ماژول‌های توجه به طور کلی عملکرد شبکه پایه را بهبود می‌بخشد؛ اما مدل پیشنهادی YOLO-CA-NET با کسب امتیاز  $mAP@0.5$  معادل ۰/۳۲۹ و میانگین بازیابی ۰/۵۱۴، برتری چشمگیری نسبت به سایر ساختارها

همان‌طور که نتایج تفکیکی مندرج در شکل ۳ نشان می‌دهد، کلاس‌های ساختاری بزرگ و واضح از نظر آناتومیک به طور مطلق بالاترین میزان دقت را کسب کرده‌اند؛ به طوری که پاتولوژی‌های کاردیومگالی و بزرگی آئورت به ترتیب به امتیازهای برجسته ۰/۶۰۱ و ۰/۵۹۴ دست یافته‌اند. از سوی دیگر، ضایعات ظریف، مینیاتوری و کم‌کنتراستی نظیر گره و توده به امتیاز ۰/۲۶۱ دست یافته‌اند که با توجه به ماهیت سخت، ابعاد مینیاتوری و پس‌زمینه درهم‌تنیده این ناهنجاری‌ها در تصاویر رادیوگرافی قفسه سینه، عملکردی قابل‌قبول محسوب می‌شود و تاییدکننده کارایی مکانیزم توجه هماهنگ در حفظ مشخصه‌های فضایی این دسته از ضایعات است. در زیربخش بعدی، جهت کالبدشکافی دقیق‌تر سهم این ماژول و توجیه فنی انتخاب آن در مقایسه با مدل پایه و سایر مکانیزم‌ها، نتایج ارزیابی حاصل از مطالعه حذفی ارائه می‌گردد.

جهت ارزیابی سهم و اثرگذاری دقیق هر یک از

جهت دار و مشخصه های فضایی ضایعات کم کنتراست قفسه سینه است که در سایر مکانیزم ها، به دلیل اتکا بر عملیات تجمیع سراسری و از دست رفتن جزئیات مکانی، مخدوش می گردیدند.

دارد. این پیکربندی توانسته است نسبت به مدل پایه YOLOv11، معیار دقت کلی را ۹/۳ درصد و میانگین بازیابی را به میزان معناداری ارتقا دهد. این امر نشان دهنده کارایی منحصر به فرد ماژول توجه هماهنگ در حفظ توأمان اطلاعات

| جدول ۴: نتایج ارزیابی حذفی لایه ها و مقایسه با سایر مکانیزم های توجه بر روی مجموعه آزمون |                             |       |             |
|------------------------------------------------------------------------------------------|-----------------------------|-------|-------------|
| معماری و پیکربندی مدل                                                                    | میانگین دقت متوسط (mAP@0.5) | دقت   | نرخ بازیابی |
| YOLOv11 (مدل پایه)                                                                       | ۰/۳۰۱                       | ۰/۳۱۲ | ۰/۴۶۱       |
| YOLOv11 + SE Block                                                                       | ۰/۳۰۸                       | ۰/۳۱۹ | ۰/۴۷۵       |
| YOLOv11 + CBAM                                                                           | ۰/۳۱۴                       | ۰/۳۲۶ | ۰/۴۹۲       |
| YOLO-CA-NET (مدل پیشنهادی)                                                               | ۰/۳۲۹                       | ۰/۳۴۱ | ۰/۵۱۴       |

نشان می دهند که مدل پایه YOLOv11 دارای ۹/۴ میلیون پارامتر و ۲۱/۵ گیگافلاپس حجم محاسبات است. با ادغام استراتژیک ماژول توجه هماهنگ در ستون فقرات شبکه و توسعه مدل YOLO-CA-NET، تعداد پارامترها تنها به ۹/۶ میلیون و حجم محاسبات به ۲۲/۱ گیگافلاپس رسیده است. این افزایش بسیار ناچیز (حدود ۰/۲ میلیون پارامتر و ۰/۶ گیگافلاپس سربار محاسباتی) نشان می دهد که مکانیزم توجه هماهنگ به دلیل ماهیت سبک و غنی سازی موازی لایه ها، بار محاسباتی سنگینی به شبکه تحمیل نمی کند.

علاوه بر ارزیابی سنجه های تشخیصی، بررسی هزینه های محاسباتی مدل ها از حیث تعداد پارامترهای یادگیری و حجم محاسبات بر حسب میلیارد عملیات ممیز شناور (GFLOPs)، شاخصی کلیدی برای سنجش قابلیت استقرار کاربردی مدل در محیط های بالینی است. جهت ارزیابی دقیق اثرگذاری ادغام ماژول توجه هماهنگ بر پیچیدگی ساختاری شبکه، مشخصات محاسباتی مدل پیشنهادی در مقایسه با مدل پایه YOLOv11 و سایر معماری ها در جدول ۵ ارائه شده است. یافته های مندرج در جدول ۵

| جدول ۵: مقایسه هزینه های محاسباتی و پیچیدگی ساختاری مدل های مختلف |                            |                      |                             |
|-------------------------------------------------------------------|----------------------------|----------------------|-----------------------------|
| معماری و مدل شبکه                                                 | تعداد پارامترها (Millions) | حجم محاسبات (GFLOPs) | میانگین دقت متوسط (mAP@0.5) |
| Faster R-CNN [۱۱،۳۰] (ResNet50-FPN)                               | ۴۱/۵                       | ۱۳۴/۰                | ۰/۲۴۱                       |
| Sparse R-CNN [۳۱]                                                 | ۴۰/۰                       | ۱۵۲/۰                | ۰/۲۴۹                       |
| YOLOv8s                                                           | ۱۱/۲                       | ۲۸/۶                 | ۰/۲۶۵                       |
| YOLOv11s (مدل پایه)                                               | ۹/۴                        | ۲۱/۵                 | ۰/۳۰۱                       |
| YOLO-CA-NET (مدل پیشنهادی)                                        | ۹/۶                        | ۲۲/۱                 | ۰/۳۲۹                       |

و ایده آلی را میان توانمندی تشخیصی بالینی و هزینه های محاسباتی برقرار ساخته است.

به منظور ارزیابی روایی متدولوژیک و اطمینان از اینکه بهبود عملکرد مدل پیشنهادی نسبت به مدل پایه دستخوش نوسانات تصادفی داده ها نبوده است، آزمون آماری

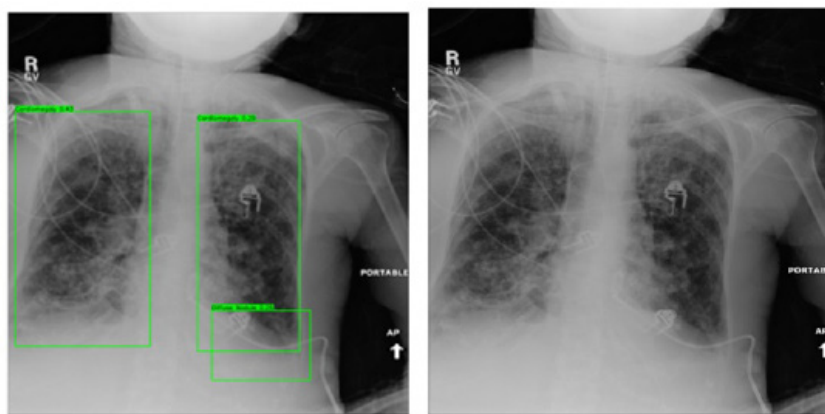
در همین حال، مدل پیشنهادی در مقایسه با آشکارسازهای دو مرحله ای مانند Faster R-CNN، افزون بر کسب دقت بسیار بالاتر (۰/۳۲۹ در برابر ۰/۲۴۱)، بیش از ۷۶ درصد کاهش در تعداد پارامترها و بیش از ۸۳ درصد کاهش در حجم محاسبات (GFLOPs) را به نمایش گذاشته است. این تحلیل اثبات می کند چارچوب پیشنهادی توازن بهینه

پیشنهادی در مواجهه با شرایط واقعی بالینی، نمونه‌ای از تصویر رادیوگرافی قفسه سینه ورودی به همراه کادرهای مرزی پیش‌بینی شده توسط مدل YOLO-CA-NET در شکل ۴ ارائه شده است.

تحلیل بصری شکل ۴ نشان‌دهنده اثربخشی معماری پیشنهادی در مواجهه با ناهنجاری‌هایی با ابعاد و ویژگی‌های بصری متفاوت است. توانایی مدل در شناسایی هم‌زمان ضایعات ساختاری بزرگ (مانند کاردیومگالی) و نودول‌های کوچک با کنتراست پایین، مؤید کارآمدی مکانیسم توجه هماهنگ در غلبه بر چالش تغییر مقیاس و حفظ ویژگی‌های محلی در لایه‌های عمیق شبکه است. علاوه بر این، دقت انطباق کادرهای مرزی با مرزهای آناتومیک ضایعات نشان می‌دهد که شبکه بدون نیاز به پیش‌پردازش‌های سنگین اضافی، قادر به تفکیک پاتولوژی‌های ریز از بافت‌های پیچیده زمینه است؛ یافته‌ای که می‌تواند در کاهش خطاهای ناشی از خستگی در فرآیند غربالگری رادیولوژیست‌ها بسیار راهگشا باشد.

بوت‌استرپ زوجی با انجام ۱۰/۰۰۰ مرتبه نمونه‌برداری مجدد بر روی مجموعه آزمون مستقل اجرا شد. تفاضل واقعی معیار میانگین دقت متوسط ( $mAP@0.5$ ) بین دو مدل برابر  $0.28 \pm$  (بهبود مطلق ۲/۸ درصدی) اندازه‌گیری شد. نتایج حاصل از تحلیل بوت‌استرپ نشان داد که بازه اطمینان ۹۵ درصد این تفاضل عملکردی در محدوده  $(0.12, 0.44)$  قرار دارد؛ از آنجا که این بازه کاملاً مثبت بوده و عدد صفر (نشان‌دهنده فرض صفر یا عدم تفاوت عملکردی) را شامل نمی‌شود، برتری معماری پیشنهادی از نظر آماری اثبات می‌گردد. همچنین، محاسبه دقیق مقدار احتمال، معناداری این ارتقا را با شاخص  $P=0.03$  تایید نمود. از آنجا که این مقدار بسیار کوچک‌تر از سطح آستانه قرارداد شده ( $P<0.05$ ) است، فرضیه صفر به طور قاطع رد شده و مشخص می‌گردد که ادغام ماژول توجه هماهنگ توانسته است به صورت کاملاً مستدل و غیرتصادفی، قدرت تشخیصی شبکه را در محلی‌سازی ناهنجاری‌های ریوی ارتقا دهد.

جهت ارزیابی دیداری و تحلیل کیفی عملکرد چارچوب



(ب) تصویر خروجی

(الف) تصویر ورودی

شکل ۴: نمونه ارزیابی کیفی خروجی مدل پیشنهادی؛ (الف) تصویر ورودی خام رادیوگرافی، (ب) تصویر خروجی مدل. کادرهای سبز رنگ نشان‌دهنده توانمندی مدل در محلی‌سازی صحیح و هم‌زمان ناهنجاری‌های ساختاری بزرگ (مانند کاردیومگالی) و ضایعات ظریف با کنتراست پایین (نظیر نودول‌های منتشر ریوی) در یک بستر آناتومیک پیچیده است.

پژوهش نشان داد که مدل پیشنهادی YOLO-CA-NET با کسب میانگین دقت ۰/۳۲۹، نه تنها نسبت به مدل پایه، بلکه در مقایسه با آشکارسازهای کلاسیک و قدرتمندی نظیر Faster R-CNN و Sparse R-CNN برتری آماری معناداری دارد. این بهبود عملکرد، به‌ویژه در افزایش ۱۹ درصدی نرخ

## بحث و نتیجه‌گیری:

هدف از این مطالعه، توسعه و ارزیابی یک مدل هوش مصنوعی کارآمد برای تشخیص خودکار ۱۴ ناهنجاری شایع در تصاویر رادیوگرافی قفسه سینه است. یافته‌های اصلی

کاربرد عملی این یافته‌ها در محیط‌های بالینی، می‌تواند به عنوان یک ابزار غربالگری اولیه برای رادیولوژیست‌ها باشد تا احتمال خطای انسانی ناشی از خستگی یا پیچیدگی تصویر کاهش یابد. با توجه به سرعت بالای استنتاج در معماری YOLOv11، این مدل قابلیت استقرار در سیستم‌های پکس بیمارستانی را داراست. این توانمندی موازنه عملیاتی و صحت عملکرد مدل در بازشناسی هم‌زمان ناهنجاری‌های فضایی و ساختاری، به طور عینی در قالب تحلیل کیفی شبیه‌سازی شده در شکل ۴ (بخش یافته‌ها) به تصویر کشیده شده است.

تحلیل تفکیکی عملکرد مدل پیشنهادی بر روی کلاس‌های مختلف نشان می‌دهد که میزان کارایی تشخیصی در پاتولوژی‌های متداول و بزرگ‌مقیاس (مانند کاردیومگالی و بزرگی آئورت) در سطح بسیار بالایی قرار دارد، اما مدل در مواجهه با برخی کلاس‌های ظریف‌تر (نظیر آتلکتازی، پنوموتوراکس و کانسایدیشن) با چالش و کاهش نرخ تشخیص مواجه است. ناکامی نسبی مدل در این لایه‌ها را می‌توان ناشی از سه عامل فنی و بالینی کلیدی دانست:

نخست، پدیده عدم تعادل شدید کلاس‌ها: همان‌طور که در توزیع آماری داده‌ها مستند شده است، کلاس‌هایی مانند آتلکتازی (۷۱٪ درصد) و پنوموتوراکس (۶۴٪ درصد) کمترین سهم را از کل بانک اطلاعاتی در اختیار دارند. در فرآیند آموزش شبکه‌های عصبی عمیق، توابع زیان به طور طبیعی تحت تأثیر کلاس‌های اکثریت قرار می‌گیرند؛ در نتیجه، گرادین‌های بهینه‌سازی به سمت الگوهای پرتکرار سوگیری پیدا کرده و مدل فرصت کافی برای یادگیری ویژگی‌های متمایزکننده کلاس‌های کم‌نمونه را پیدا نمی‌کند.

دوم، ظرافت ساختاری و ابعاد مینیاتوری ضایعات: پاتولوژی‌هایی مانند گره‌های ریوی یا لایه‌های نازک هوا در پنوموتوراکس، بخش بسیار کوچکی از ماتریس تصویر (۶۴۰\*۶۴۰) را اشغال می‌کنند. در معماری‌های مبتنی بر CNN، با عبور تصویر از لایه‌های متوالی کانولوشنی و اعمال فرآیندهای نمونه‌کاهی، ویژگی‌های فضایی ظریف

بازیابی، تایید می‌کند که تعبیه صریح اطلاعات موقعیتی در فرآیند یادگیری ویژگی، برای شناسایی پاتولوژی‌های پیچیده ریوی بسیار حیاتی است.

توصیف دقیق‌تر نتایج نشان می‌دهد که مکانیزم توجه هماهنگ با ایجاد آگاهی فضایی در شبکه، توانسته است چالش اصلی تصاویر CXR یعنی هم‌پوشانی ساختارهای آناتومیک (مانند دنده‌ها و عروق) با ضایعات ظریف را تا حد زیادی مرتفع کند. برخلاف مدل‌های متداول که در فرآیند استخراج ویژگی، جزئیات مربوط به اشیاء کوچک و ظریف را از دست می‌دهند، مدل پیشنهادی با حفظ مختصات دقیق فضایی، مانع از محو شدن ویژگی‌های پاتولوژیک ظریف در لایه‌های عمیق شبکه می‌شود. برتری عملکرد مدل در کلاس‌هایی مانند «توده» و «گره» مویید این مطلب است که مدل توانسته است وابستگی‌های فضایی دوربرد را بهتر درک کرده و ناهنجاری‌های کوچک را از بافت‌های زمینه‌ای متمایز نماید. این یافته با نتایج مطالعات قبلی که بر اهمیت مکانیزم‌های توجه در تصویربرداری پزشکی تاکید داشتند همسو است [۳۳، ۳۲، ۱۸]؛ با این تمایز که مدل پیشنهادی ما توانست بدون تحمیل هزینه محاسباتی سنگین مدل‌های ترانسفورمر، به دقتی فراتر از مدل‌های دو مرحله‌ای سنگین دست یابد.

از نقاط قوت این مطالعه می‌توان به استفاده از یکی از بزرگ‌ترین و معتبرترین مجموعه‌داده‌های عمومی (VinDr-CXR) و پیاده‌سازی بر روی سخت‌افزار پیشرفته اشاره کرد که امکان تعمیم‌پذیری نتایج را افزایش می‌دهد. با این حال، مطالعه حاضر با محدودیت‌هایی نیز روبرو بود؛ از جمله تمرکز بر یک مجموعه‌داده واحد و وجود عدم تعادل شدید در تعداد نمونه‌ها در کلاس‌های ظریفی مانند آتلکتازی بود. نتایج نشان داد که حتی در این کلاس‌های دشوار و ظریف نیز، مدل YOLO-CA-NET نسبت به YOLOv11 استاندارد، نرخ تشخیص مثبت را به طور معناداری ارتقا داده است که نشان از کارایی مکانیزم توجه در تمرکز بر نواحی مشکوک و کوچک دارد. یافته پایداری بالای مدل در تشخیص ناهنجاری‌های ساختاری (مانند کاردیومگالی) علی‌رغم تغییرات شدید کنتراست، نشان‌دهنده اثربخشی پیش‌پردازش CLAHE در کنار معماری پیشنهادی است.

ویژگی‌های جمعیتی خاص، فاقد تنوع نژادی و بومی لازم برای تعمیم‌پذیری مطلق در سطح بین‌المللی است. در این راستا، تأیید نهایی این مدل در محیط‌های بالینی داخل کشور نیازمند ارزیابی بر روی داده‌های محلی است. این موضوع به عنوان فاز دوم پژوهش تعریف شده است تا با جمع‌آوری و حاشیه‌نویسی دیتاست‌های بومی قابلیت تعمیم‌پذیری مدل در سطح ملی ارتقا یابد. همچنین وزن‌های آموزش‌دیده این مدل به‌راحتی از طریق فرآیند یادگیری انتقالی روی داده‌های بومی بیمارستان‌های کشور قابل استفاده است. در پژوهش‌های آتی، تمرکز بر جمع‌آوری داده‌های چندمرکزی، پیاده‌سازی آزمون‌های بالینی میدانی جهت سنجش نرخ توافق پزشک-ماشین، و به‌کارگیری استراتژی‌های اعتبارسنجی خارجی جهت تضمین پایداری و تعمیم‌پذیری مدل پیشنهادی در اولویت قرار خواهد گرفت

### نتیجه‌گیری:

این پژوهش با معرفی چارچوب YOLO-CA-NET، گامی در جهت بهینه‌سازی مدل‌های تشخیص شیء برای کاربردهای پزشکی برداشت. نتایج نشان داد که ادغام استراتژیک مکانیزم توجه سبک و آگاه از فضا در ستون فقرات شبکه، جهت‌گیری بسیار مؤثری برای توسعه ابزارهای تشخیصی دقیق و در عین حال سریع است. در مجموع، مدل پیشنهادی توازن بهینه‌ای بین دقت تشخیصی و هزینه محاسباتی ارائه داد که برای کاربردهای دنیای واقعی هوش مصنوعی در سلامت ضروری است.

همچنین در راستای گسترش یافته‌های پژوهش حاضر و عبور از محدودیت‌های متدولوژیک موجود، چندین مسیر راهبردی برای تحقیقات آتی پیشنهاد می‌گردد:

نخست، به منظور ارزیابی تعمیم‌پذیری فرامرزی و پایداری مدل، پیشنهاد می‌شود معماری ارائه شده در قالب اعتبارسنجی خارجی بر روی مجموعه‌داده‌های چندمرکزی و متنوع‌تر نظیر MIMIC-CXR مورد سنجش قرار گیرد. افزون بر این، برای بهره‌گیری از بانک‌های اطلاعاتی کلان فاقد کادر مرزی (مانند CheXpert و ChestX-ray14)، استفاده از این مراجع در قالب استراتژی‌های یادگیری

این اشیاء کوچک در نقشه‌های ویژگی لایه‌های عمیق محو یا مستهلک می‌شوند. اگرچه مازول توجه هماهنگ با حفظ اطلاعات جهت‌دار توانسته است این نقیصه را تا حدودی جبران کند، اما چالش آشکارسازی اشیاء مینیاتوری در رادیوگرافی همچنان پابرجا است.

سوم، شباهت بالای درون‌کلاسی و تداخل پاتولوژیک: از منظر بالینی، تظاهرات بصری پاتولوژی‌هایی چون کانسالی‌دیشن، ارتشاح ریوی و کدورت ریوی در تصاویر رادیوگرافی قفسه سینه شباهت‌های ساختاری و فازی بسیار شدیدی با یکدیگر دارند. این هم‌پوشانی بالای ویژگی‌های بصری سبب می‌شود که مدل در شرایط کمبود داده‌های آموزشی برای این کلاس‌های خاص، در تفکیک مرزهای پاتولوژیک دچار خطا شده و نرخ مثبت کاذب یا منفی کاذب در آن‌ها افزایش یابد.

با وجود دستاوردهای ارزشمند معماری پیشنهادی YOLO-CA-NET در ارتقای قدرت تشخیصی و محلی‌سازی ناهنجاری‌های ریوی، این مطالعه مواجه با محدودیت‌هایی است که توجه به آن‌ها، مسیر را برای توسعه کاربردی در آینده هموار می‌سازد.

نخست، از منظر پروتکل‌های یادگیری عمیق، اگرچه داده‌ها با نسبت استاندارد تفکیک شدند و فرآیند اعتبارسنجی به صورت درون‌برنامه‌ای در حین آموزش صورت گرفت، اما این پژوهش فاقد یک مجموعه اعتبارسنجی کاملاً مستقل و خارجی جهت ارزیابی ثانویه پیش از فاز آزمون نهایی بوده است؛ ساختاری که می‌تواند تا حدودی بر ارزیابی سوگیری‌های پنهان مدل اثرگذار باشد.

دوم، ارزیابی کارایی چارچوب پیشنهادی صرفاً بر پایه سنجش‌های محاسباتی رایانه‌ای (mAP، دقت و بازیابی) بر روی مراجع آرشیوی انجام شده و مدل حاضر فاقد ارزیابی‌های بالینی واقعی در محیط‌های زنده بیمارستانی و سنجش در چرخه کاری پزشکان است.

سوم، دیتاست VinDr-CXR که مبنای آموزش مدل بوده، به دلیل تمرکز بر پروتکل‌های تصویربرداری و

مقاله، تایید نهایی نسخه انتشاری، مسئولیت پاسخگویی در مورد صحت کار؛ حمیدرضا غفاری: مشارکت در تجزیه و تحلیل و تفسیر داده‌ها، بازنگری و مرور انتقادی محتوا، تایید نهایی، مسئولیت پاسخگویی.

### تعارض منافع

نویسندگان این مقاله اعلام می‌دارند که هیچ‌گونه وابستگی سازمانی یا تضاد منافع (اعم از مالی یا شخصی) در رابطه با موضوع پژوهش و نتایج ارائه شده در این مقاله وجود ندارد.

### تشکر و قدردانی

از دانشگاه آزاد اسلامی واحد فردوس بابت همکاری در این پژوهش که برگرفته از رساله دکترای تخصصی می‌باشد، قدردانی می‌گردد.

انتقالی جهت پیش‌آموزش شبکه یا به‌کارگیری آن‌ها در چارچوب‌های آشکارسازی کم‌نظارت پتانسیل بالایی برای توسعه مدل خواهد داشت.

دوم، با توجه به چالش عدم تعادل شدید کلاس‌ها در پاتولوژی‌های نادر، بررسی و توسعه توابع زیان تخصصی (مانند Focal Loss توازن یافته یا Dice Loss) می‌تواند از مسیرهای جذاب برای بهبود نرخ موازنه گرادین‌ها باشد. همچنین، به منظور ارتقای اعتمادپذیری بالینی و پذیرش مدل توسط رادیولوژیست‌ها، پیاده‌سازی متدهای هوش مصنوعی قابل تفسیر نظیر الگوریتم‌های Grad-CAM++ جهت نقاشی نقشه‌های حرارتی مبانی تصمیم‌گیری مدل در لایه‌های عمیق در دستور کار قرار دارد.

سوم، توسعه مدل‌های ترکیبی و تلفیقی نوین از طریق ادغام توانمندی استخراج ویژگی‌های موضعی خانواده YOLO با قدرت مدل‌سازی وابستگی‌های محیطی بلندمدت در معماری‌های مبتنی بر ویژن ترانسفورمرها در لایه‌های نهایی شبکه، افق‌های جدیدی را در غنی‌سازی هم‌زمان مشخصه‌های فضایی و کانالی به‌ویژه در شکار پاتولوژی‌های مینیاتوری و ظریف خواهد گشود.

### ملاحظات اخلاقی:

#### پیروی از اصول اخلاق در پژوهش

این مطالعه دارای تاییدیه اخلاقی به شماره IR.IAU.BIRJAND.REC.1404.036 از دانشگاه آزاد اسلامی واحد فردوس است.

### حامی مالی

این پژوهش به صورت مستقل انجام شده و هیچ‌گونه حمایت مالی از سوی سازمان‌های دولتی، خصوصی یا غیرانتفاعی دریافت نکرده است.

### مشارکت نویسندگان

ملیکا قاسمی: مشارکت در مفهوم‌آفرینی و طراحی مطالعه، مشارکت در گردآوری، نرم‌افزار، نگارش پیش‌نویس

## References

- World Health Organization. To X-ray or not to X-ray?. World Health Organization; 2016. [Link](#)
- Smith-Bindman R, Kwan ML, Marlow EC, Theis MK, Bolch W, Cheng SY, et al. Trends in use of medical imaging in US health care systems and in Ontario, Canada, 2000-2016. *JAMA*. 2019; 322(9):843-56. DOI: [10.1001/jama.2019.11456](#) PMID: [31479136](#)
- Mustafa Qafesha R, Daa Hindawi M, Elbadry M, Sharabati I, Mashaly D, Dervis M, et al. Chest radiography in interstitial lung disease: Accuracy and radiological features from a systematic review and meta-analysis. *Sarcoidosis Vasc Diffuse Lung Dis*. 2025; 42(2):16108. DOI: [10.36141/svdl.v42i2.16108](#) PMID: [40699492](#)
- Ahmad HK, Milne MR, Buchlak QD, Ektas N, Sander-son G, Chamtie H, et al. Machine learning augmented interpretation of chest X-rays: A systematic review. *Diagnostics* (Basel). 2023; 13(4):743. DOI: [10.3390/diagnostics13040743](#) PMID: [36832231](#)
- Lian J, Liu J, Zhang S, Gao K, Liu X, Zhang D, et al. A structure-aware relation network for thoracic diseases detection and segmentation. *IEEE Trans Med Imaging*. 2021; 40(8):2042-52. DOI: [10.1109/TMI.2021.3070847](#) PMID: [33819152](#)
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. 2014. DOI: [10.48550/arXiv.1409.1556](#)
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. CVPR. 2016; 770-78. [Link](#)
- Iandola F, Moskewicz M, Karayev S, Girshick R, Darrell T, Keutzer K. Densenet: Implementing efficient convnet descriptor pyramids. *arXiv preprint arXiv:1404.1869*. 2014. DOI: [10.48550/arXiv.1404.1869](#)
- Balakrishnan T, Sandeep SS. Enhanced GELAN with VGG-STYLE ConvNets for brain tumour detection. *arXiv preprint 2024 [Internet]*. *arXiv preprint arXiv:2405.03541*, 2024. DOI: [10.48550/arXiv.2405.03541](#)
- Suzuki K. Overview of deep learning in medical imaging. *Radiol Phys Technol*. 2017; 10(3):257-73. DOI: [10.1007/s12194-017-0406-5](#) PMID: [28689314](#)
- Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*. 2017; 39(6):1137-49. DOI: [10.1109/TPAMI.2016.2577031](#) PMID: [27295650](#)
- Hidayatullah P, Syakrani N, Sholahuddin MR, Gelar T, Tubagus R. YOLOv8 to YOLO11: A comprehensive architecture in-depth comparative review. *arXiv preprint arXiv:2501.13400*. 2025. DOI: [10.29207/resti.v10i2.6598](#)
- Mao M, Hong M. YOLO object detection for real-time fabric defect inspection in the textile industry: A review of YOLOv1 to YOLOv11. *Sensors* (Basel). 2025; 25(7):2270. DOI: [10.3390/s25072270](#) PMID: [40218782](#)
- Nguyen HQ, Lam K, Le LT, Pham HH, Tran DQ, Nguyen DB, et al. VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations. *Sci Data*. 2022; 9(1):429. DOI: [10.1038/s41597-022-01498-w](#) PMID: [35858929](#)
- Soydaner D. Attention mechanism in neural networks: where it comes and where it goes. *Neural Comput Appl*. 2022; 34(16):13371-85. DOI: [10.1007/s00521-022-07366-3](#)
- Woo S, Park J, Lee JY, Kweon IS. Cbam: Convolutional block attention module. In *European conference on computer vision 2018* (pp. 3-19). Cham: Springer International Publishing. DOI: [10.1007/978-3-030-01234-2\\_1](#)
- Mahendran N. Squeeze aggregated excitation network. *arXiv preprint arXiv:2308.13343*. 2023. DOI: [10.48550/arXiv.2308.13343](#)
- Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. In *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* 2021 (pp. 13708-13717). IEEE. DOI: [10.1109/CVPR46437.2021.01350](#)
- Yan J, Zeng Y, Lin J, Pei Z, Fan J, Fang C, et al. Enhanced object detection in pediatric bronchoscopy images using YOLO-based algorithms with CBAM attention mechanism. *Heliyon*. 2024; 10(12):e32678. DOI: [10.1016/j.heliyon.2024.e32678](#) PMID: [39021922](#)
- Zhou N, Xu M, Shen B, Hou K, Liu S, Sheng H, et al. Vit-unet: A vision transformer based unet model for coastal wetland classification based on high spatial resolution imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2024; 17:19575-87. DOI: [10.1109/JSTARS.2024.3487250](#)
- Shehzadi T, Hashmi KA, Liwicki M, Stricker D, Afzal MZ. Object detection with transformers: A review. *Sensors* (Basel). 2025; 25(19):6025. DOI: [10.3390/s25196025](#) PMID: [41094848](#)
- Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, Shah M. Transformers in vision: A survey. *ACM computing surveys (CSUR)*. 2022; 54(10s):1-41. DOI: [10.1145/350524](#)
- Khanam R, Hussain M. Yolov11: An overview of the



- key architectural enhancements. arXiv preprint arXiv:2410.17725. 2024. DOI: [10.48550/arXiv.2410.17725](https://doi.org/10.48550/arXiv.2410.17725)
24. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In Proceedings of the AAAI conference on artificial intelligence. 2019; 33(1):590-97. DOI: [10.48550/arXiv.1901.07031](https://doi.org/10.48550/arXiv.1901.07031)
  25. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2017 (pp. 2097-2106). DOI: [10.48550/arXiv.1705.02315](https://doi.org/10.48550/arXiv.1705.02315)
  26. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems. 2019; 32. [Link](#)
  27. Jocher G, Qiu J. Ultralytics YOLO11 software and documentation. 2024. [Link](#)
  28. Jocher, G, Qiu J. Ultralytics Yolo11. GitHub: San Francisco, CA, USA 2024. [Link](#)
  29. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. IEEE Trans Pattern Anal Mach Intell. 2020; 42(8):2011-23. DOI: [10.1109/TPAMI.2019.2913372](https://doi.org/10.1109/TPAMI.2019.2913372) PMID: [31034408](https://pubmed.ncbi.nlm.nih.gov/31034408/)
  30. Team PT. Torchvision models documentation: Faster R-CNN ResNet-50 FPN. PyTorch Documentation. 2023. [Link](#)
  31. Sun P, Zhang R, Jiang Y, Kong T, Xu C, Zhan W, et al. Sparse R-CNN: An end-to-end framework for object detection. IEEE Trans Pattern Anal Mach Intell. 2023; 45(12):15650-64. DOI: [10.1109/TPAMI.2023.3292030](https://doi.org/10.1109/TPAMI.2023.3292030) PMID: [37402189](https://pubmed.ncbi.nlm.nih.gov/37402189/)
  32. Tang P, Yan X, Hu X, Wu K, Lasser T, Shi K. Anomaly detection in medical images using encoder-attention-2decoders reconstruction. IEEE Trans Med Imaging. 2025; 44(8):3370-82. DOI: [10.1109/TMI.2025.3563482](https://doi.org/10.1109/TMI.2025.3563482) PMID: [40293916](https://pubmed.ncbi.nlm.nih.gov/40293916/)
  33. Dai W, Liu R, Wu Z, Wu T, Wang M, Zhou J, et al. Exploiting scale-variant attention for segmenting small medical objects. IEEE Trans Neural Netw Learn Syst. 2026; 37(7):3161-75. DOI: [10.1109/TNNLS.2025.3645355](https://doi.org/10.1109/TNNLS.2025.3645355) PMID: [41525526](https://pubmed.ncbi.nlm.nih.gov/41525526/)