Applied Computer Systems
ISSN 2255-8691 (online)
2026, vol. 31, no. 1, pp. 41–53
https://doi.org/10.2478/acss-2026-0004
https://content.sciendo.com

1862
RIGA TECHNICAL UNIVERSITY

Sensitivity-Oriented YOLOv11 for Robust Multi-Label Lesion Detection in Chest X-rays

Thi-Da-Huong Truong[1], Ngoc Huynh Pham[2], Vo-Phuong-Tam Nguyen[3], Thi-Thanh-Thuy Le[4], Hai Thanh Nguyen[5*]
[1–5]*College of Information and Communication Technology, Can Tho University, Can Tho, Vietnam*

***Abstract* – Chest X-ray lesion detection remains challenging due to severe class imbalance, subtle lesion appearance, and the risk of over-optimistic evaluation caused by improper data splitting. In this study, we propose a sensitivity-oriented detection framework based on YOLOv11 for robust chest X-ray screening under clinically realistic conditions. The proposed approach integrates patient-wise data partitioning, enhanced data augmentation, and prediction fusion to improve generalization while mitigating data leakage. Experiments are conducted on the VinDr-CXR dataset using a strict patient-level split to ensure full separation between training and validation sets. A series of internal fine-tuning scenarios is designed to analyse the trade-offs among precision, recall, and localization accuracy. Based on internal validation, the medium-scale YOLOv11-m configuration (denoted as M3) is selected as the reference model, as it provides the most stable balance between sensitivity and localization performance. Under rigorous evaluation, M3 achieves a precision of 0.431, a recall of 0.416, an mAP@0.5 of 0.387, and an mAP@0.5:0.95 of 0.193. Compared with representative baselines, M3 demonstrates improved robustness under patient-wise evaluation, outperforming transformer-based DETR by a large margin (mAP@0.5: 0.387 vs. 0.232) and achieving performance comparable to YOLOv7 while exhibiting substantially higher sensitivity to small and diffuse lesions. Further comparison with recent studies shows that the proposed method achieves higher overall mAP@0.5 (0.387 vs. 0.362–0.378) while improving detection performance on clinically challenging abnormality classes. These results indicate that the proposed YOLOv11-based framework provides a reliable and clinically meaningful baseline for chest X-ray lesion screening and future methodological advancements.**

***Keywords* – lesion detection, screening, transformer-based, YOLO.**

I. Introduction

Chest X-ray imaging is among the most widely used and cost-effective diagnostic modalities in clinical practice, serving a critical role in screening and assessing a broad spectrum of thoracic abnormalities, such as cardiomegaly, pulmonary nodules, consolidations, and pleural diseases [1], [2]. Its accessibility and low radiation dose make chest X-ray a routine first-line imaging tool in large-scale screening programs and emergency settings. Nevertheless, interpretation of chest X-ray images remains inherently challenging, even for experienced radiologists, due to overlapping anatomical structures, subtle lesion appearances, and substantial inter-class variability.

Recent advances in deep learning-based object detection frameworks have enabled promising automation of chest X-ray lesion detection and localization [3], [4]. One-stage detectors, in particular, have gained attention for their balance between detection accuracy and computational efficiency, making them suitable for real-time and large-scale screening. The YOLO family of models is widely adopted in medical imaging due to its simplicity, rapid inference, and competitive localization performance. Despite these developments, chest X-ray lesion detection remains challenging. A primary obstacle is the severe class imbalance in publicly available datasets, where a few common abnormalities dominate the training distribution, while rare or subtle lesions are underrepresented [5]–[8]. This imbalance often results in biased learning, with models achieving high precision for dominant classes but poor sensitivity to clinically critical, underrepresented abnormalities. In screening contexts, missed detections of such lesions may pose greater clinical risk than false alarms, highlighting the necessity for sensitivity-oriented detection strategies. Additionally, the evaluation protocols employed in many studies represent a critical, yet frequently overlooked, issue.

Many chest X-ray detection studies employ image-level random splitting, which permits images from the same patient to appear in both training and evaluation sets [9], [10]. This practice introduces data leakage, resulting in over-optimistic performance estimates that do not accurately represent real-world generalization to unseen patients. As a result, reported gains in accuracy or mAP may not correspond to reliable clinical performance. Additionally, transformer-based detection frameworks have been investigated for chest X-ray analysis due to their capacity to model global contextual relationships [11]. However, these models generally require large-scale, balanced training data to fully leverage self-attention mechanisms. In the data-limited and imbalanced context of medical imaging, such approaches often struggle to converge or generalize, thereby limiting their practical utility in routine screening.

Motivated by these challenges, this study investigates a sensitivity-oriented chest X-ray lesion-detection framework based on a modern one-stage detection architecture. Instead of prioritising peak benchmark performance under permissive evaluation protocols, the proposed approach emphasises robust

* Corresponding author's e-mail: nthai.cit@ctu.edu.vn
Article received 2026-01-17; accepted 2026-03-09

and clinically meaningful detection under strict patient-wise evaluation. Through a system-level design that integrates data partitioning, training strategies, and prediction behaviour, the framework aims to establish a reliable baseline for chest X-ray screening and future methodological advancements of this study are summarised as follows:

- A clinically realistic end-to-end detection pipeline for chest X-ray lesion analysis is established, encompassing medical image preprocessing, high-resolution normalization, and strict patient-wise data partitioning to eliminate data leakage and ensure reliable generalization evaluation.
- A sensitivity-oriented adaptation of a modern YOLOv11-based detector is proposed for chest X-ray screening, demonstrating that contemporary one-stage architectures can effectively handle severe class imbalance and subtle lesion patterns under rigorous evaluation settings.
- A label refinement strategy based on Weighted Box Fusion (WBF) is integrated to address annotation noise arising from multi-reader bounding boxes, enabling more stable training and improved localization consistency on complex medical datasets.
- A comprehensive, unbiased experimental evaluation protocol is conducted, including internal validation, baseline comparison, and benchmarking against recent studies using consistent patient-level splits, providing meaningful insights into model behaviour across lesion types and clinical risk profiles.

II. Related Work

A. CNN-based Lesion Detection on Chest X-rays

Early deep learning approaches for chest X-ray (CXR) analysis mainly focused on abnormality classification using convolutional neural networks (CNNs). Multi-label and multi-class classification models based on DenseNet and EfficientNet reported AUC values above 0.80 on large-scale datasets such as ChestX-ray14 and related benchmarks [12]–[14]. Although effective for image-level diagnosis, these methods do not provide explicit lesion localization, which limits their clinical interpretability.

Fully supervised object detection frameworks were later introduced to address lesion localization. Pham et al. [15] proposed an ensemble of Faster R-CNN, YOLOv5, and EfficientDet optimised by Weighted Box Fusion (WBF), achieving an mAP@0.4 of 0.292 on VinDr-CXR. Similarly, Nguyen et al. [16] demonstrated that incorporating lung-region extraction before YOLO-based detection slightly improved AP@0.5 and AP@0.5:0.95 on VinDr-CXR. Despite their accuracy, two-stage or ensemble-based detectors often incur high computational cost and limited inference speed, constraining their use in real-time screening.

One-stage detectors offer a more favourable balance between efficiency and performance. RetinaNet-based approaches with transfer learning achieved mAP@0.5 values up to 0.55 for selected disease subsets on VinDr-CXR [17]. Recent YOLO-based systems further improved detection speed and accuracy; Al-antari et al. [18] reported detection accuracy exceeding 96 % for COVID-19-related abnormalities, while Mustafa and Nsour [19] achieved mAP@0.5 values around 0.83–0.85 using YOLOv8. However, several studies indicate that recall for small or low-contrast lesions often remains below 0.40, suggesting limited sensitivity in clinically challenging cases [15], [16].

B. Weakly-supervised and Unsupervised Abnormality Localization

To reduce annotation costs, weakly supervised and unsupervised localization methods have been explored. Yu et al. [20] proposed an anatomy-guided weakly-supervised frame- work, achieving recall values of 0.58–0.74 at low IoU thresholds (IoU@0.1–0.25) on MIMIC-CXR, but performance degraded significantly at stricter localization criteria. Unsupervised anomaly detection approaches, such as α-GAN-based reconstruction models [21] and self-supervised multiresolution patch learning [22], reported AUROC values of 0.73–0.75 on the RSNA and PadChest datasets. More recently, Sheng et al. [23] employed cross-domain self-supervised learning and improved mAP@0.5 to 0.289 on VinDr-CXR, yet this remains lower than that of fully supervised detectors. Overall, weakly-supervised and unsupervised methods exhibit inferior localization precision and sensitivity compared to bounding-box–supervised detectors, limiting their suitability for high- reliability screening applications.

C. Transformer-based and Hybrid Detection Frameworks

Hybrid detection frameworks combining CNNs with additional modalities or architectural components have also been investigated. Hsieh et al. [24] introduced MDF-Net, which fuses chest X-rays with clinical data, and achieved AP = 31.69 % on the MIMIC-Eye dataset, outperforming Mask R-CNN at the expense of increased model complexity and reliance on non-imaging inputs. Although transformer-based paradigms are gaining attention, current evidence suggests that their benefits in chest X-ray detection remain limited under severe class imbalance and small lesion conditions. Hybrid designs typically yield modest improvements of approximately 1–2 % AP while increasing computational overhead [24], making them less attractive for large-scale, real-time screening compared to optimised CNN-based one-stage detectors.

D. Evaluation Protocols, Data Leakage, and Clinical Reliability

Beyond model design, the evaluation protocol critically affects reported performance. The VinDr-CXR dataset was clinically validated by Nguyen et al. [25], who emphasised strict patient-level separation to avoid data leakage. Behrendt et al. [26] further demonstrated that improper image-level splitting can lead to inflated detection metrics, with enforced patient-wise splits reducing mAP by 5–10 % and recall by up to 12 %. Importantly, most prior works prioritise mAP optimisation, while sensitivity is often underreported. Clinical screening studies highlight that false negatives pose substantially higher risk than false positives, particularly for subtle or early-stage abnormalities [12], [27]. Consequently, recent YOLO-based studies have begun to emphasise recall-

oriented configurations, as demonstrated by Xie et al. [27], who achieved recall values above 0.94 using a YOLOv11-derived architecture for pneumonia detection.

In summary, existing chest X-ray lesion-detection methods achieve competitive accuracy in fully supervised settings but often overestimate real-world performance due to permissive evaluation protocols and limited emphasis on sensitivity. Weakly-supervised and hybrid approaches remain constrained by localization accuracy and complexity. Motivated by these limitations, this study proposes a sensitivity-oriented chest X-ray lesion detection framework based on YOLOv11, evaluated under strict patient-wise data partitioning. The proposed methodology aims to deliver robust generalization and clinically realistic screening performance, as detailed in the following section.

III. Proposed Method

A. Overall System Pipeline

To address the problem of detecting and classifying abnormalities in chest X-ray images – which is challenged by small lesion sizes, high morphological variability, low contrast, and label noise arising from multi-expert annotations – this study proposes an end-to-end medical object detection system designed to simultaneously improve localization accuracy, training stability, and practical applicability. The core idea of the proposed approach is to decompose the overall task into a set of complementary functional modules, each targeting a specific challenge inherent to chest X-ray data. These modules include input data standardization and enrichment, label noise reduction induced by multi-radiologist annotations, multi-scale feature learning for small and rare lesions, and automated inference in a practical deployment environment. Rather than being treated as isolated steps, these components are integrated into a unified pipeline, ensuring consistency throughout both training and inference stages.

Figure 1 illustrates the overall architecture and data flow of the proposed system. The system takes chest X-ray images as input, performs image conversion and normalization, and simultaneously fuses bounding box annotations from multiple radiologists using a weighted fusion mechanism to generate consensus ground truth. Based on the cleaned dataset, a detection model trained on the YOLOv11 architecture learns multi-scale representations, particularly well suited for small and subtle abnormalities. During inference, the trained model directly processes unseen chest X-ray images and outputs predicted bounding boxes, corresponding abnormality labels, and confidence scores. The entire system operates as a closed-loop automated pipeline and can be seamlessly integrated into a web-based interface to demonstrate its applicability in real-world clinical scenarios. The detailed design of each system component is described in the following subsections.

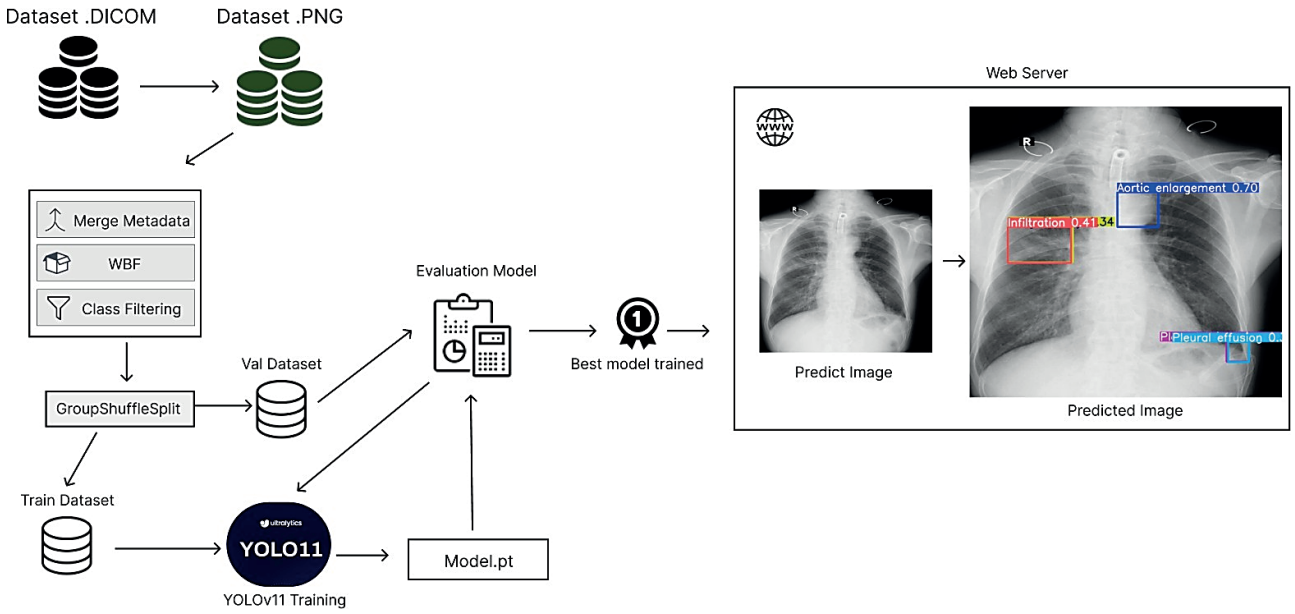


Fig. 1. The overall framework for multi-label lesion detection in chest x-rays.

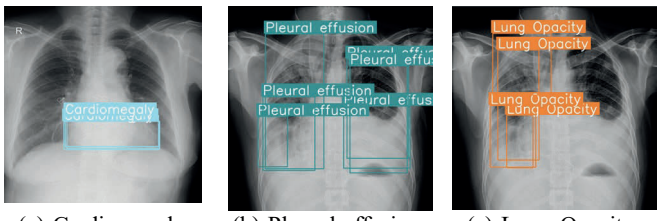
B. Dataset and Preprocessing

Dataset. This study uses the VinBigData Chest X-ray Abnormalities Detection (VinDr-CXR) dataset [28], a publicly available benchmark for chest X-ray abnormality detection constructed from real-world clinical data in Vietnam. The dataset comprises approximately 18 000 posteroanterior (PA) chest X-ray images in DICOM format, collected from multiple hospitals and representing diverse patient characteristics. 17 experienced radiologists provided annotations. In the training set, each image was independently annotated by three radiologists, and in the test set, five radiologists reviewed each image, with final labels determined by expert consensus. Each image may contain zero or multiple abnormalities, annotated using bounding boxes and corresponding labels. The dataset includes 14 abnormality categories (Table I), along with a "No finding" class indicating the absence of detectable pathology. As illustrated in Fig. 2, the VinDr-CXR dataset covers a broad

spectrum of abnormality scales, from prominent and salient findings such as cardiomegaly to small and subtle lesions including nodules and calcifications. VinDr-CXR is adopted due to its large scale, rigorous multi-expert annotation protocol, and widespread adoption in international benchmarks, making it a reliable dataset for evaluating chest X-ray abnormality detection models.

TABLE I
ABNORMALITY CATEGORIES IN THE VINDR-CXR DATASET

Class ID	Abnormality	Class ID	Abnormality
0	Aortic enlargement	7	Lung Opacity
1	Atelectasis	8	Nodule/Mass
2	Calcification	9	Other lesion
3	Cardiomegaly	10	Pleural effusion
4	Consolidation	11	Pleural thickening
5	ILD (Interstitial Lung Disease)	12	Pneumothorax
6	Infiltration	13	Pulmonary fibrosis



(a) Cardiomegaly (b) Pleural effusion (c) Lung Opacity

(d) Nodule/Mass (e) Calcification (f) Pneumothorax

Fig. 2. Representative chest X-ray abnormalities from the VinDr-CXR dataset. The first row shows common abnormalities with relatively clear visual patterns, while the second row illustrates small-scale or infrequent lesions that are more challenging to detect.

Figure 3 highlights the severe class imbalance, where No Finding dominates the dataset, while several clinically meaningful abnormalities appear with substantially fewer samples. This imbalance motivates the use of targeted augmentation and robust training strategies, as discussed in Section III-B3.

The bounding box area distribution (Fig. 4) reveals extreme scale variability both across and within abnormality classes. Several classes exhibit heavy-tailed distributions with numerous outliers, reflecting the heterogeneous manifestation of chest pathologies in clinical practice. This observation highlights the need for multi-scale feature learning and motivates the use of high input resolution and advanced augmentation strategies.

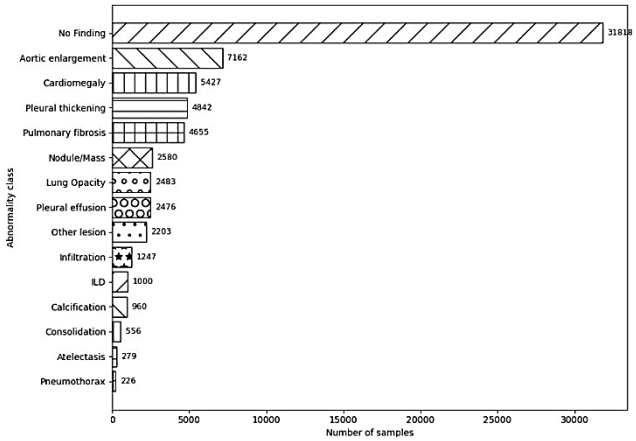


Fig. 3. Horizontal bar chart showing the distribution of abnormality classes in the VinDr-CXR dataset.

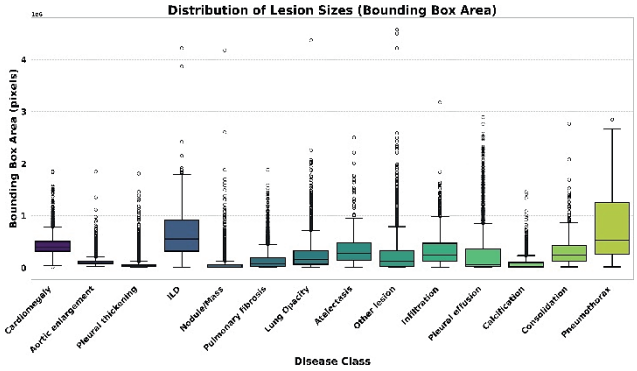


Fig. 4. Distribution of bounding box area across abnormality classes.

Image Preprocessing and Normalization. Chest X-ray images are preprocessed to ensure spatial and intensity consistency before being fed into the deep learning model. Specifically, the original DICOM images are converted to standard image formats (PNG/JPEG) while preserving grayscale intensity information critical for diagnosis. All images are resized to a unified resolution of 1024 × 1024 to retain fine-grained details of small and subtle lesions and to maintain consistency during training. In addition, intensity normalization is applied to mitigate variations caused by different imaging devices and acquisition conditions across data sources. Images lacking clear diagnostic information are excluded to improve overall data quality.

Data Augmentation. As a prerequisite for data augmentation and model training, all annotations and bounding boxes are carefully preprocessed, including coordinate transformations to match the resized input resolution and label normalization to conform to the target model format (e.g., YOLO-style annotations). To enhance generalization and address severe class imbalance, strong data augmentation strategies are employed during training. Advanced techniques such as Mosaic [26], which combines four images into one, and MixUp [29], which linearly blends image–label pairs, are applied to increase sample diversity and training difficulty. Furthermore, Copy-Paste augmentation [30] is utilised to enrich rare abnormality classes (e.g., Nodule and Calcification) by inserting lesion regions into other images. In addition, standard geometric and

photometric augmentations, such as horizontal flipping, rotation, and brightness adjustment, are optionally applied to improve model robustness further.

Label Processing and Noise Reduction. Due to the multi-expert annotation protocol of VinDr-CXR, a single abnormality may be represented by multiple overlapping bounding boxes with varying localization. Directly using these raw annotations can introduce label noise and degrade model performance. To address this issue, Weighted Box Fusion (WBF) [23] is applied to merge overlapping boxes of the same class into a single consensus annotation. Compared with Non-Maximum Suppression, WBF preserves spatial information through weighted averaging, which is more suitable for medical images with ambiguous lesion boundaries.

After fusion, bounding boxes are converted into the target detection format (e.g., YOLO-style annotations) by normalizing coordinates with respect to the input image resolution. This process reduces inter-observer variability and provides more stable ground-truth labels for model training.

C. Proposed Detection Model

In this study, YOLOv11 is adopted as the primary detection model for chest X-ray abnormality detection and serves as the main framework for comparison with baseline methods. YOLOv11 is a recent evolution of the YOLO family, designed to balance detection accuracy and inference efficiency while supporting effective fine-tuning on domain-specific datasets, such as medical images. The model follows the standard Backbone-Neck-Head paradigm and incorporates improved building blocks, including C3k2 and SPPF, to enhance multi-scale feature representation (Fig. 5). In addition, the C2PSA (Parallel Spatial Attention) module is integrated to improve spatial attention, particularly useful for capturing subtle and low-contrast abnormalities commonly found in chest X-ray images. YOLOv11 is initialised from COCO-pretrained weights using official Ultralytics implementations (e.g., YOLOv11-s or YOLOv11-m) and fine-tuned on the VinDr-CXR dataset.

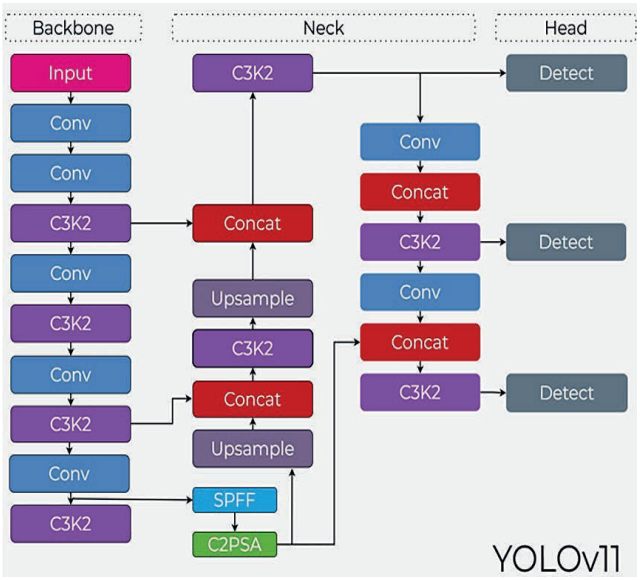


Fig. 5. An illustration of YOLOv11 architecture [31].

D. Training Strategy

The training process is conducted in multiple stages (M0–M3) to improve model robustness and detection accuracy progressively. Each stage explores different optimisation strategies, data augmentation strengths, and YOLOv11 model scales. The detailed configurations of each training stage are summarised in Table II. Early stages (M0–M1) focus on baseline optimisation using SGD with moderate augmentation. Later stages (M2–M3) adopt AdamW, more substantial data augmentation, and Weighted Box Fusion to better handle annotation noise and class imbalance. Different YOLOv11 variants are evaluated under the same protocol to analyse the trade-off between model capacity and detection performance.

TABLE II
TRAINING CONFIGURATIONS ACROSS DIFFERENT EXPERIMENTAL STAGES

Stage	Optimiser	Lr_0	Lr_f	MixUp	CutMix	Degree	Mosaic	Model	WBF
M0	SGD	0.0005	0.001	0.01	0.01	0.0	1.0	YOLOv11-l	–
M1	SGD	0.0005	0.001	0.20	0.15	0.0	1.0	YOLOv11-l	–
M2	AdamW	0.0010	0.010	0.15	0.00	10.0	1.0	YOLOv11-s	Yes
M3	AdamW	0.0010	0.010	0.15	0.00	10.0	1.0	YOLOv11-m	Yes

E. Implementation Details

All experiments are conducted on the VinDr-CXR dataset with all images resized to 1024 × 1024 to preserve small lesion details. The dataset is partitioned into training and validation sets using patient-wise GroupShuffleSplit (Image_id) to avoid data leakage, with a split ratio of 85:15 for M0–M1 and 80:20 for M2–M3. The later models (M2–M3) adopt a slightly larger validation set to enable more stable and reliable performance evaluation during model comparison Models are trained for 100 epochs with a batch size of 8 or 16, depending on GPU memory. A Cosine Annealing scheduler is applied to improve convergence stability. All experiments are implemented in Python 3.11 using the Ultralytics framework, and executed on NVIDIA Tesla T4 GPUs (16 GB) in Kaggle and Google Colab environments.

IV. Experimental Results and Analysis

A. An Overview of Experimental Scenarios

The experimental evaluation is conducted under multiple complementary scenarios to provide a comprehensive and unbiased assessment of the proposed approach:

- *Internal fine-tuning scenarios (M0–M3)*: A sequence of controlled training configurations is designed to investigate the effects of optimisation strategies, data augmentation intensity, and YOLOv11 model scales under a unified patient-wise data split and evaluation protocol.
- *Baseline model comparison*: YOLOv7 [32] is adopted as a representative one-stage detector, while DETR [33] represents transformer-based object detection with global self-attention. Both baseline models are re-trained or fine-tuned using publicly available implementations and evaluated under the same patient-level data partitioning and metric settings as the proposed YOLOv11-based model, enabling a controlled and fair comparison on chest X-ray images.
- *Comparison with previous studies*: The proposed method is further compared with representative approaches reported in prior chest X-ray lesion detection studies to position its performance within the existing literature under comparable evaluation settings.
- *Visual and practical analysis*: Qualitative inspection of detection outputs and confidence-based behaviour is conducted to assess robustness, interpretability, and practical reliability in realistic clinical screening scenarios.

B. Evaluation Metrics

In this study, the selected evaluation metrics were used to comprehensively assess the performance of models for detecting lesions in X-ray images, including both classification and localization. Specifically, precision [34] and recall [34] measure prediction accuracy and the ability to detect lesions, F1-Score [35] balances the two, while mAP@0.5 [36] and mAP@0.5:.95 [37] reflect how well predicted bounding boxes match the ground truth across different thresholds. The Confusion Matrix [38] is used to analyse misclassifications between lesion classes visually.

Loss [39] functions monitored during training evaluate convergence and improve the model's localization and classification capabilities. In summary, this set of evaluation metrics not only measures predictive performance but also ensures the model operates effectively and reliably in clinical applications.

C. Internal Validation: Fine-tuning Scenarios

Overall Performance across M0–M3. The results on the internal validation set (Fig. 6) indicate that the M0–M3 training scenarios achieve stable performance under the same evaluation protocol, clearly reflecting the trade-offs among precision, recall, and spatial localization accuracy. Across all scenarios, the models obtain precision values within a narrow range (0.416–0.458), suggesting a relatively conservative and consistent prediction behaviour. M1 achieves the highest precision, but at the cost of lower recall, indicating a strategy that prioritises reducing false positives. In contrast, M2 and M3 exhibit higher recall, with M3 achieving the highest value (0.416), demonstrating an improved ability to detect lesions on previously unseen data. This characteristic is fundamental in medical screening applications, where minimising false negatives is critical.

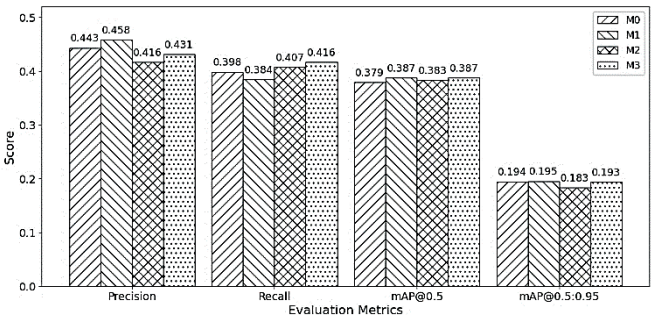
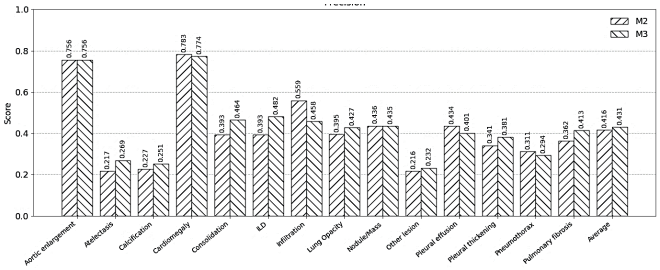


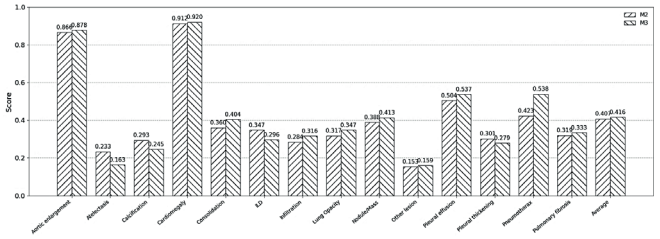
Fig. 6. Overall performance comparison of fine-tuning scenarios M0–M3 on the validation set.

Regarding mAP@0.5, the four scenarios yield comparable results, indicating similar overall detection performance at a moderate IoU threshold. However, under the more stringent metric mAP@0.5:0.95, the early-stage models (M0–M1) slightly outperform the later ones, while M2 shows a noticeable decline. This trend highlights the trade-off between aggressive data augmentation strategies and precise bounding-box regression. Overall, the results indicate that M3 provides the best balance among precision, recall, and mAP, while maintaining a sensitivity-oriented behaviour. Based on this internal validation, M3 is selected as the reference configuration for all subsequent in-depth analyses and comparisons with external and state-of-the-art methods.

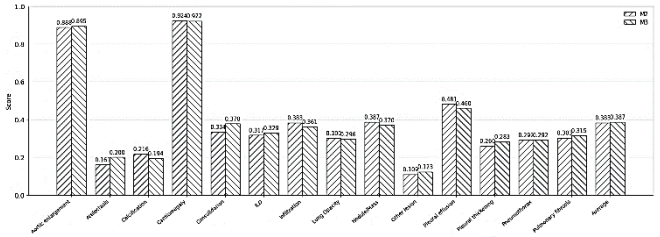
Detailed Class-wise Performance (M2 vs M3). Figure 7 presents a class-wise comparison between the M2 and M3 models, where evaluation metrics are selected according to the morphological characteristics of each lesion group, for large-structure lesions with well-defined boundaries, such as Cardiomegaly and Aortic enlargement, spatial localization quality is critical. In this group, M3 shows a clear advantage in terms of AP@0.5:0.95. Notably, the Cardiomegaly class achieves an AP@0.5:0.95 of 0.635 with M3, compared to 0.561 with M2, indicating more accurate bounding-box regression when using the medium-scale architecture. For small and localized lesions, including nodules/masses and calcifications, recall is a key metric to reduce missed detections. The results show that recall for the Nodule/Mass class increases from 0.388 (M2) to 0.413 (M3), suggesting that M3 is more effective at capturing fine-grained features, which is particularly important in early clinical screening.



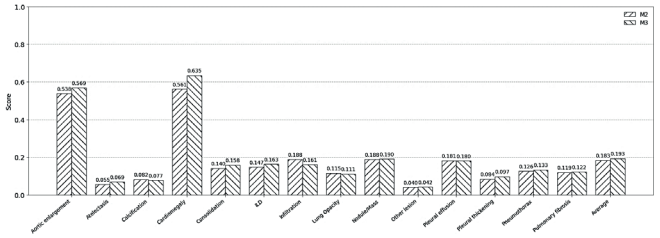
(a) Precision



(b) Recall



(c) AP@0.5

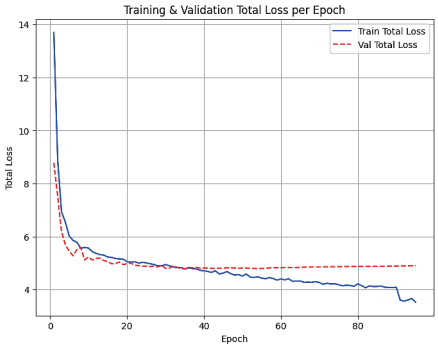


(d) AP@0.5:0.95

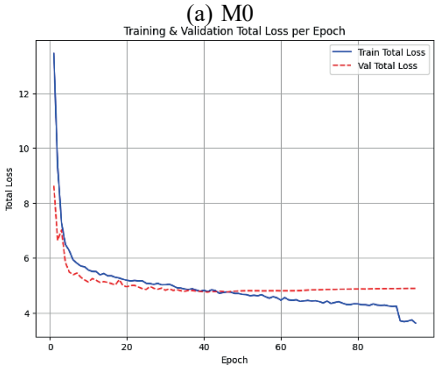
Fig. 7. Class-wise comparison between M2 and M3 on the validation set under different evaluation metrics.

For diffuse lesions with ambiguous boundaries, such as Consolidation and Infiltration, the two models exhibit different trends. M3 achieves a higher AP@0.5 for the Consolidation class, whereas M2 performs slightly better for Infiltration. This observation indicates that, for highly diffuse patterns, the performance difference between M2 and M3 remains limited, suggesting that increased model capacity does not necessarily translate into consistent gains for such lesion types. For rare and challenging classes, including Atelectasis, Other lesion, and Pleural thickening, M3 yields only marginal improvements in precision and AP@0.5:0.95. The limited performance gap reflects the inherent difficulty posed by class imbalance and heterogeneous lesion appearances. Overall, the class-wise analysis indicates that the advantages of M3 primarily stem from improved localization accuracy and greater sensitivity to clinically significant lesions. In contrast, performance on certain diffuse lesion classes remains comparable between the two models.

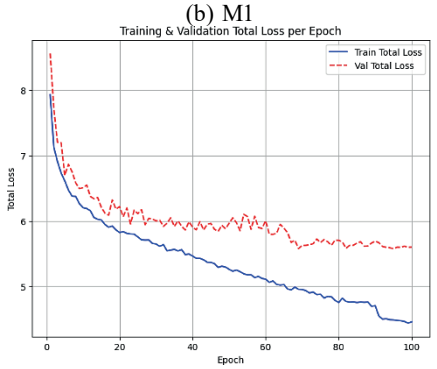
Visual and Practical Analysis. Figure 8 highlights a clear distinction between the early-trained models (M0–M1) and the enhanced models (M2–M3). M0 and M1 exhibit early loss saturation, where the training loss continues to decrease faster than the validation loss, indicating mild overfitting. This behaviour is consistent with the quantitative results, as these models achieve higher precision but lower recall, reflecting limited generalization to unseen data. In contrast, M2 and M3 maintain higher, yet more stable, validation losses throughout training.



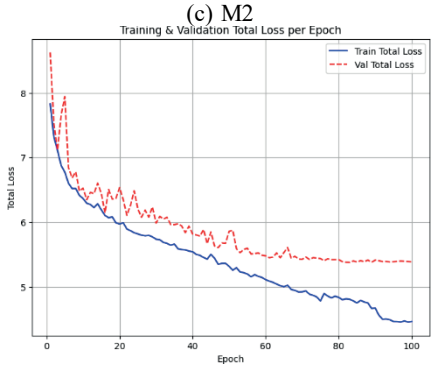
(a) M0



(b) M1



(c) M2



(d) M3

Fig. 8. Training and validation loss curves of internal fine-tuning scenarios M0–M3.

This slower but consistent convergence corresponds to improvements in average recall from 0.407 to 0.416 and mAP@0.5:0.95 from 0.183 to 0.193, indicating that data augmentation and prediction fusion strategies enable the model to learn more generalizable features, particularly for rare and challenging lesions.

Figures 9–11 jointly demonstrate the stability and robustness of the M3 model during training. Training and validation losses decrease smoothly across epochs without late-stage divergence, indicating stable convergence and no noticeable overfitting. While precision and recall converge to balanced values of 0.431 and 0.416, respectively, mAP@0.5 and mAP@0.5:0.95 steadily increase and saturate at 0.387 and 0.193, respectively. Furthermore, the narrow interquartile ranges observed in the boxplots further confirm that performance is consistently maintained across epochs rather than driven by isolated peaks. Together, these results indicate reliable convergence and stable generalization, supporting the suitability of M3 for practical deployment.

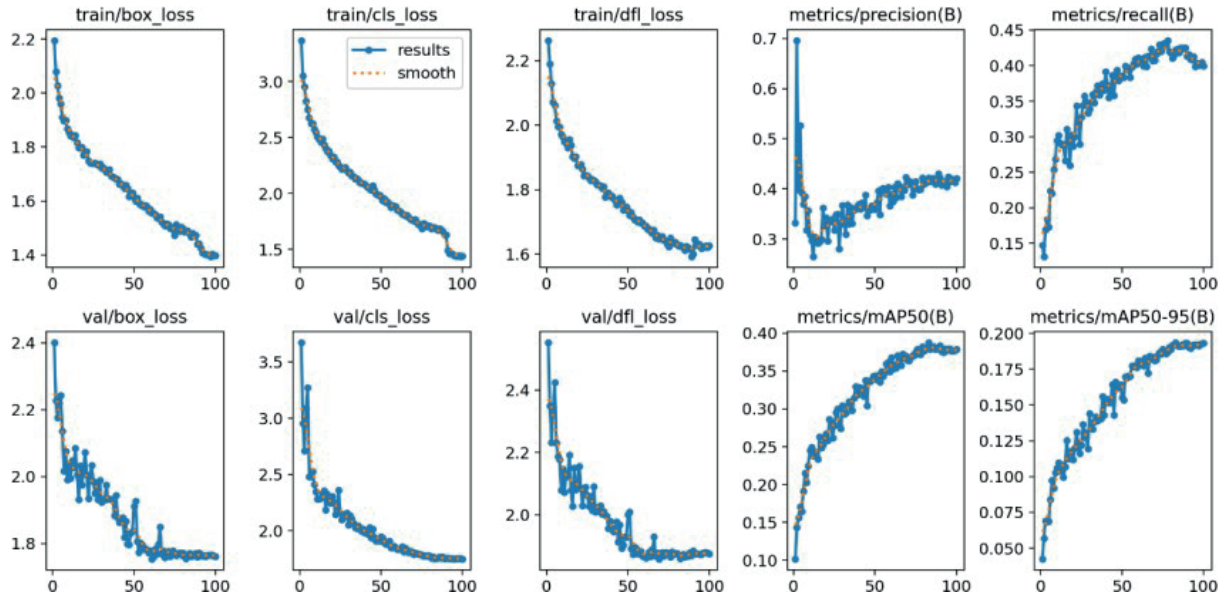


Fig. 9. Training and validation loss curves of the proposed M3 model.

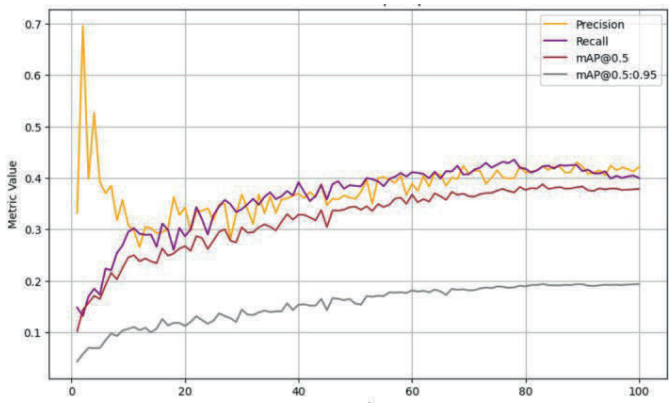


Fig. 10. Evolution of evaluation metrics over training epochs for the M3 model.

The raw confusion matrix (Fig. 12) reveals a severe class imbalance in the ground-truth distribution. Common abnormalities, such as Aortic enlargement and Cardiomegaly, dominate the true-positive counts (561 and 408 cases, respectively), whereas rare classes, such as Atelectasis (2 true positives) and Infiltration (9 true positives), provide insufficient samples for robust feature learning. This imbalance directly explains the low precision observed in rare classes, which remains in the range of 0.216–0.269. A large number of false negatives is observed in the background row, particularly for diffuse and low-contrast lesions. For example, Pleural thickening records 684 missed cases compared to only 108 correct detections, while Pulmonary fibrosis exhibits 534 missed cases. These statistics indicate that the M3 model adopts a conservative prediction strategy, prioritising precision over recall for visually ambiguous abnormalities.

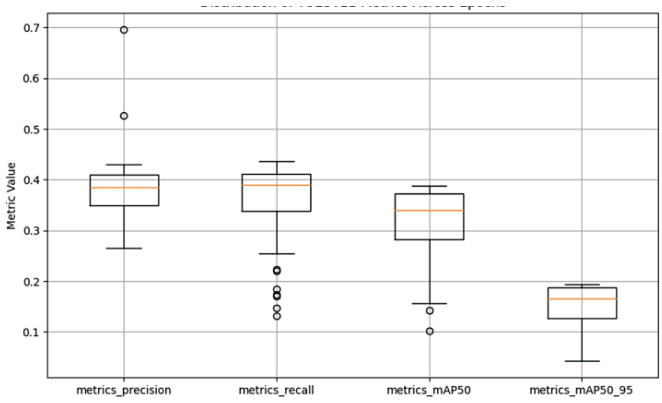


Fig. 11. Boxplot distribution of performance metrics across training epochs for the M3 model.

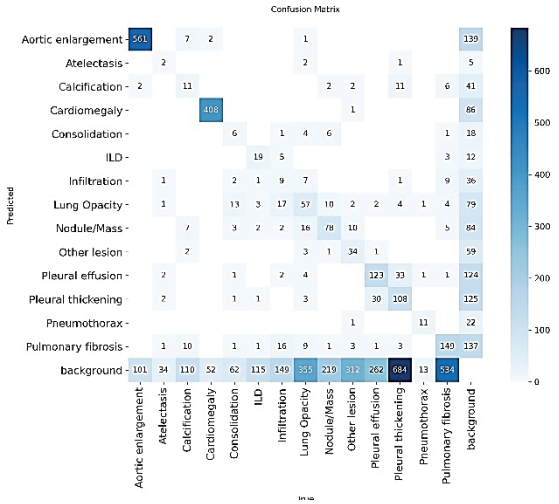


Fig. 12. Raw confusion matrix of the M3 model on the validation set.

The normalized confusion matrix (Fig. 13) further highlights class-dependent variations in recall. Lesions with well-defined anatomical structures achieve high sensitivity, such as Cardiomegaly (recall = 0.920) and Aortic enlargement (recall = 0.878), whereas small or subtle abnormalities, including Atelectasis, remain challenging (recall = 0.163). In addition, notable cross-class confusion is observed among visually similar pulmonary patterns, with Lung Opacity frequently misclassified as Consolidation or Infiltration. This reflects the inherent difficulty of distinguishing diffuse parenchymal abnormalities on 2D chest X-ray images. Overall, these results confirm that the M3 model is well suited for screening prominent structural abnormalities. At the same time, detecting rare and diffuse pulmonary lesions remains a challenging area for future improvement.

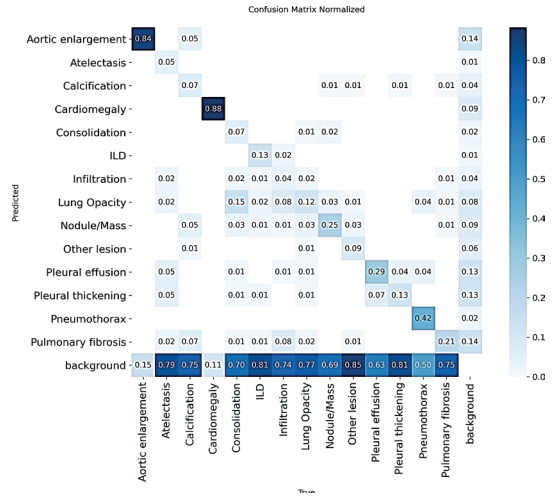


Fig. 13. Normalized confusion matrix of the M3 model on the validation set.

The visual comparison between the ground truth annotations (Fig. 14) and the predictions generated by M3 (Fig. 15) shows that the model accurately detects and localizes large, clinically significant abnormalities. Structural findings such as Cardiomegaly and Aortic enlargement are identified with bounding boxes closely matching expert annotations and high confidence scores (typical ≥ 0.8), confirming the model's applicability in screening scenarios. In contrast, small or subtle lesions, including Nodule/Mass and Infiltration, remain challenging, with some cases being missed or detected at relatively low confidence levels (0.3–0.4). This limitation is primarily attributable to the ambiguous visual characteristics of such abnormalities on 2D chest X-ray images and the scarcity of training samples for rare classes.

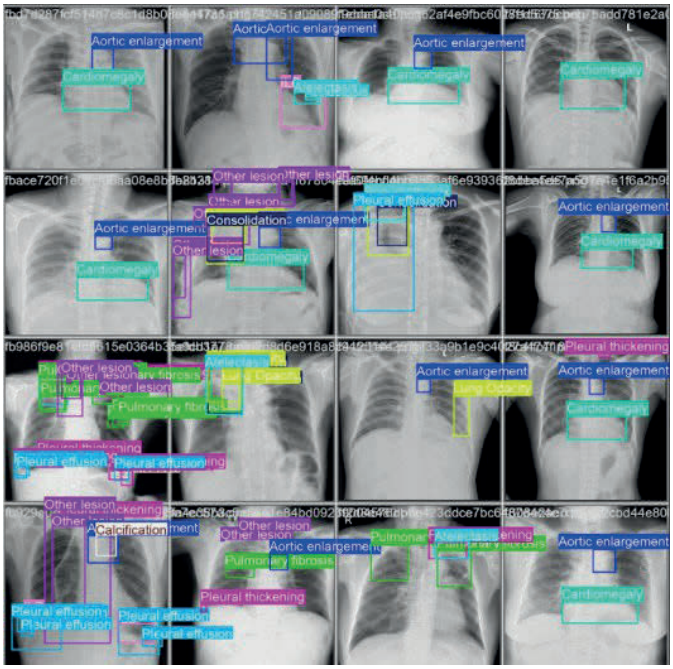


Fig. 14. Ground-truth annotations on representative validation images.

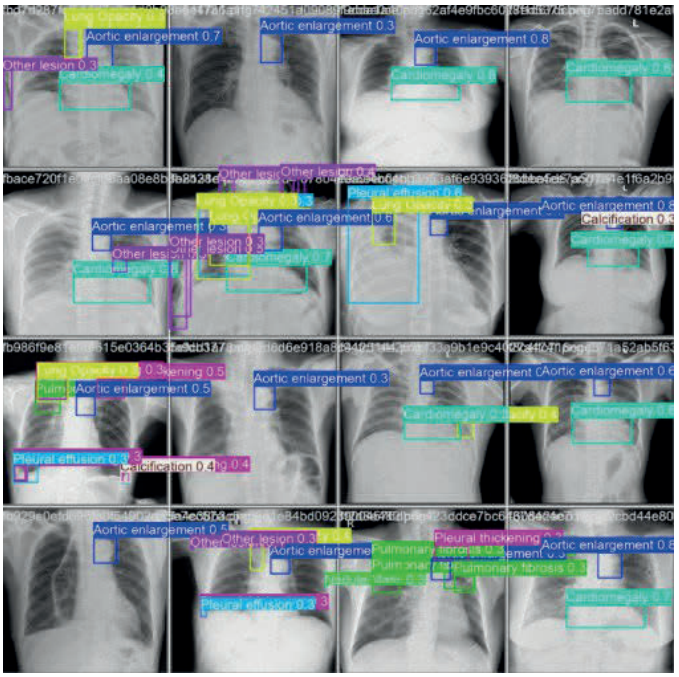


Fig. 15. Detection results of the M3 model on the same validation images.

Figure 16 indicates that the F1-score reaches its maximum value of approximately 0.41 at a confidence threshold of around

0.16, corresponding to a recall of 0.416 and a precision of 0.431. Increasing the threshold beyond 0.2 leads to a rapid decline in F1-score for rare classes, thereby increasing the risk of false negatives, whereas classes with prominent structural patterns maintain stable performance across a broader range of confidence values. Overall, selecting a low operating confidence threshold (appropriately 0.16) is appropriate for real-world deployment, allowing M3 to prioritise clinical sensitivity and function effectively as a screening tool, while accepting a moderate increase in false positives to minimise missed pathological findings. This operating point is therefore adopted in all subsequent baseline comparisons.

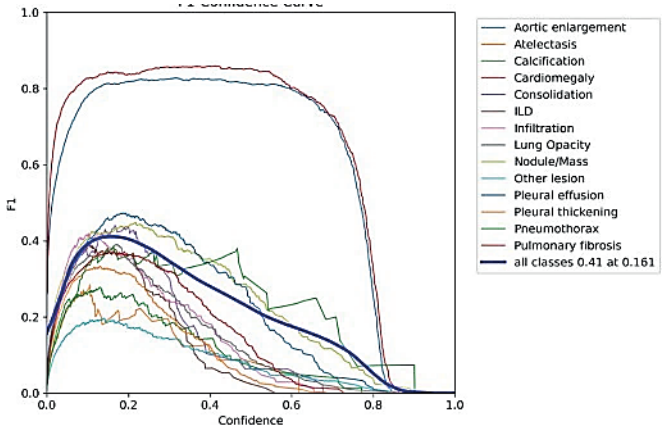


Fig. 16. F1–Confidence.

D. Baselines Comparison

Based on internal validation, M3 is selected as the most stable and well-balanced configuration and is therefore used as the reference model for baseline comparisons. Unless otherwise stated, all baselines are evaluated using the same patient-wise data split (GroupShuffleSplit) and identical evaluation metrics. As illustrated in Fig. 17, although the YOLOv11 model reported by Salah achieves a high mAP@0.5 of 0.645, this result is primarily affected by data leakage caused by image-level random splitting, where images from the same patient appear in both training and validation sets. In contrast, M3 is evaluated under a strict patient-wise split, ensuring an independent test set; thus, despite a lower mAP@0.5 of 0.387, its performance is more reliable and clinically meaningful.

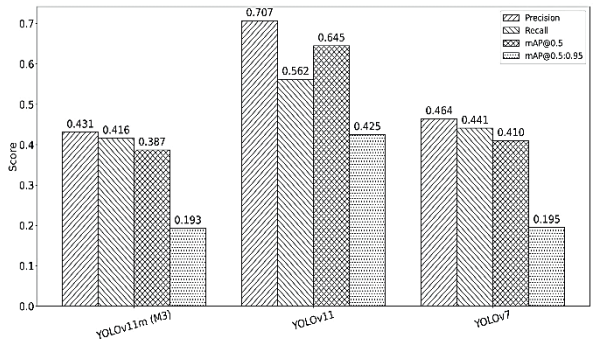


Fig. 17. Quantitative comparison of the proposed M3 model with baseline methods on the internal validation set.

Compared with YOLOv7, M3 yields a slightly lower mAP@0.5 (0.387 vs. 0.41); however, confusion matrix analysis reveals that YOLOv7 misses a substantial proportion of small and diffuse lesions, highlighting the limitations of older architectures on highly imbalanced chest X-ray data. By leveraging the YOLOv11 architecture together with enhanced data augmentation and prediction fusion, M3 improves sensitivity to challenging classes while maintaining stable localization. When compared with DETR, M3 shows a clear advantage. DETR achieves a mAP of only 0.232 even under a lower IoU threshold (0.4), whereas M3 reaches 0.387 at IoU 0.5, confirming the suitability of the proposed YOLOv11-based framework for chest X-ray lesion screening. Overall, M3 provides a balanced trade-off between precision (0.431) and recall (0.416) under rigorous patient-wise evaluation, making it a more clinically trustworthy baseline despite not achieving the highest nominal mAP.

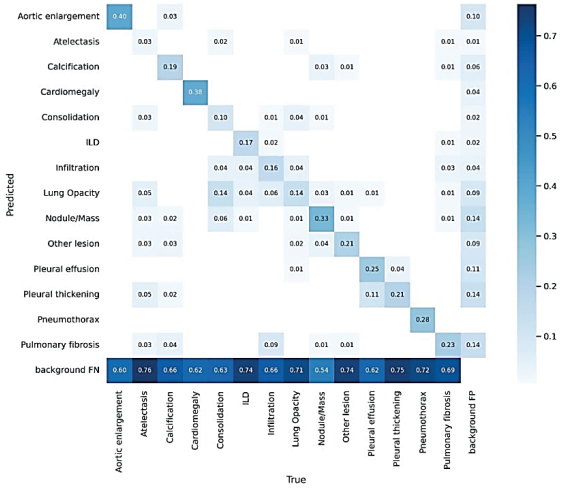


Fig. 18. Confusion matrix of the YOLOv7 model.

E. Comparison with Previous Studies

The per-class comparison indicates that M3 achieves the best overall performance (Figs. 19 and 20), with an mAP@0.5 of 0.387, outperforming DAMGNet (0.362) and ResNet-34 + YOLOv5 [16] (0.378). This result confirms the effectiveness of the YOLOv11 architecture combined with the proposed training strategy and prediction fusion scheme. Notably, the advantage of M3 is most pronounced for small- scale or visually subtle lesions, such as Nodule/Mass (0.370), Calcification (0.194), and Pneumothorax (0.292), demonstrating an improved sensitivity to challenging abnormalities that are often missed by previous approaches. For large structural abnormalities with well-defined morphology, including Cardiomegaly and Aortic enlargement, M3 maintains high performance (0.922 and 0.895, respectively), only slightly lower than methods optimized for dominant classes or deeper backbones. This observation indicates that M3 does not sacrifice performance on common, easier classes to improve performance on difficult ones, but instead achieves balanced performance across lesion groups. In contrast, for diffuse abnormalities with ambiguous boundaries, such as Consolidation and Lung Opacity, M3 yields lower scores

compared to some reference methods. This trend reflects the inherent difficulty of distinguishing diffuse parenchymal patterns on 2D chest X-ray images and suggests that no single model can consistently dominate across all lesion types.

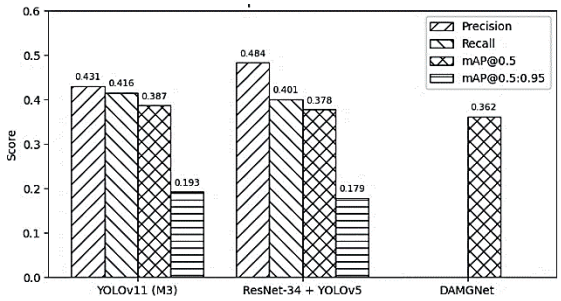


Fig. 19. Overall comparison with related studies.

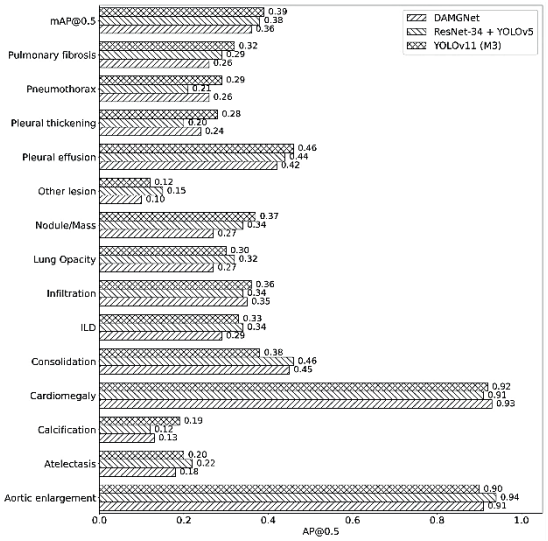


Fig. 20. Per-class comparison with related studies.

However, it should also be noted that differences in data partitioning strategies across studies may influence direct quantitative comparisons, as some previous works adopted different train–validation–test splits or did not clearly report their data division protocols.

Overall, these results demonstrate that M3 provides a well-balanced trade-off between precision and sensitivity across all classes, with particularly notable improvements for small and rare lesions. Such characteristics are critical in clinical screening settings, indicating that M3 not only remains competitive with prior studies on quantitative metrics but also offers greater practical value for early detection and decision support.

V. Discussion

This study investigates chest X-ray abnormality detection under clinically realistic screening conditions, where the primary requirement is robust generalization to unseen patients rather than maximising absolute performance metrics. To this end, a patient-wise data partitioning strategy is adopted to eliminate data leakage and ensure a fair and realistic evaluation protocol. Under this stringent setting, the achieved performance is lower than that reported in studies relying on image-level random splits. However, this discrepancy does not indicate inferior modelling capability. Instead, it highlights the gap between academic evaluation and real-world clinical deployment, where models must learn pathology-related features rather than patient-specific visual patterns.

An important insight from this work is that optimising mAP alone does not necessarily lead to effective clinical screening. In chest X-ray interpretation, false negatives pose substantially higher clinical risk than false positives. Consequently, the sensitivity-oriented training strategy adopted in the M3 configuration is particularly desirable, as it prioritises the detection of subtle and atypical abnormalities, even at the cost of increased false alarms. From a methodological perspective, the results suggest that CNN-based one-stage detectors, when combined with appropriate training strategies and prediction fusion, remain well-suited for medical imaging tasks with limited and imbalanced data. In contrast, transformer-based detectors struggle to fully exploit their modelling capacity under such constraints, primarily because they rely on large-scale, balanced training corpora.

Despite its favourable stability and generalization behaviour, the proposed approach remains constrained by the inherent limitations of two-dimensional chest X-ray imaging, especially for diffuse abnormalities with ambiguous boundaries. This observation suggests that future improvements require more advanced strategies, such as multimodal learning or semi-supervised frameworks, to better capture subtle pathological cues.

In general, the proposed method demonstrates that reliability, generalization, and clinical relevance are more critical than peak performance metrics. These characteristics position the model as a practical and dependable baseline for chest X-ray screening and a solid foundation for future developments in computer-aided diagnosis.

VI. Conclusion

This paper presents a robust chest X-ray lesion detection framework based on the YOLOv11 architecture, with a particular focus on clinically realistic evaluation and generalization to unseen patients. Unlike many prior studies that rely on image-level random splits, the proposed approach adopts a strict patient-wise data partitioning strategy to eliminate data leakage and provide a more faithful assessment of real-world screening performance. Experimental results demonstrate that the selected M3 configuration achieves a stable and well-balanced trade-off between precision and sensitivity under rigorous evaluation conditions. While its absolute mAP is lower than some previously reported results, this behaviour reflects the impact of enforcing genuine generalization rather than memorization of patient-specific patterns. In particular, the sensitivity-oriented training strategy enables improved detection of small, rare, and visually subtle lesions, which is critical in chest X-ray screening scenarios where missed abnormalities pose a higher clinical risk than false alarms. Rather than aiming to maximise benchmark

metrics alone, the proposed framework deliberately targets a practical operating point aligned with clinical screening requirements, prioritising robustness, consistency, and interpretability. A comparative analysis further indicates that modern CNN-based one-stage detectors, when combined with appropriate training strategies and prediction fusion, remain more effective than transformer- based alternatives under data-limited, highly imbalanced medical imaging conditions.

Despite these advantages, the method is constrained by intrinsic challenges of two-dimensional chest X-ray analysis, including class imbalance, overlapping anatomical structures, and ambiguous lesion boundaries. Future work will therefore explore more advanced strategies, such as multi-modal learning, semi-supervised training with large-scale unlabelled data, and uncertainty-aware prediction, to further enhance sensitivity and robustness for diffuse and rare abnormalities. Overall, this study positions the proposed YOLOv11-based framework not as a replacement for all existing detection approaches, but as a reliable and clinically meaningful baseline for patient-level chest X-ray screening, offering a solid foundation for future research toward deployable computer-aided diagnosis systems.

Declarations

Availability of Data and Material

The data supporting this study's findings are available from the study of [16].

Competing Interests

All authors declare that they have no conflicts of interest.

Authors' Contributions

T.-D.-H.T.: Data curation, Methodology, Validation, Visualization, Software, Writing – original draft, Writing – review and editing; N.H.P.: Formal analysis, Methodology, Validation, Visualization, Writing – original draft; V.-P.-T.N.: Data curation, Methodology, Validation, Visualization, Software, Writing – original draft, Writing – review and editing; T.-T.-T.L.: Data curation, Methodology, Validation, Visualization, Software, Writing – original draft, Writing – review and editing; H.T.N.: Conceptualization, Formal analysis, Methodology, Validation, Visualization, Supervision, Writing – original draft, Writing – review and editing.

Ethical Approval

The data analysed in this study were collected from publicly accessible sources and were treated as secondary data.

References

[1] A. Ait Nasser and M. A. Akhloufi, "A review of recent advances in deep learning models for chest disease detection using radiography," *Diagnostics*, vol. 13, no. 1, Jan. 2023, Art. no. 159. https://doi.org/10.3390/diagnostics13010159

[2] A. Rehman, A. Khan, G. Fatima, S. Naz, and I. Razzak, "Review on chest pathogies detection systems using deep learning techniques," *Artificial Intelligence Review*, vol. 56, pp. 12607–12653, Mar. 2023. https://doi.org/10.1007/s10462-023-10457-9

[3] Y. Han, "Comparative analysis of two-stage and one-stage object detection models," in *Proceedings of the 2nd International Conference on Data Analysis and Machine Learning*, vol. 1, Kuala Lumpur, Malaysia, 2024, pp. 289–294. https://doi.org/10.5220/0013515900004619

[4] Y. Xie, B. Zhu, Y. Jiang, B. Zhao, and H. Yu, "Diagnosis of pneumonia from chest X-ray images using yolo deep learning," *Frontiers in Neurorobotics*, vol. 19, Apr. 2025, Art. no. 1576438. https://doi.org/10.3389/fnbot.2025.1576438

[5] R. Walsh and M. Tardy, "A comparison of techniques for class imbalance in deep learning classification of breast cancer," *Diagnostics*, vol. 13, no. 1, Dec. 2022, Art. no. 67. https://doi.org/10.3390/diagnostics13010067

[6] E. Yanar, F. Kutan, K. Ayturan, U. Kutbay, O. Algın, F. Hardalaç, and A. M. Ağıldere, "A comparative analysis of the Mamba, Transformer, and CNN architectures for multi-label chest X-ray anomaly detection in the NIH chestX-ray14 dataset," *Diagnostics*, vol. 15, no. 17, Sept. 2025, Art. no. 2215. https://doi.org/10.3390/diagnostics15172215

[7] V.-T.-N. Pham, Q.-C. Nguyen, and Q.-V. Nguyen, "Chest X-rays abnormalities localization and classification using an ensemble framework of deep convolutional neural networks," *Vietnam Journal of Computer Science*, vol. 10, no. 1, pp. 55–73, Aug. 2022. https://doi.org/10.1142/S2196888822500348

[8] F. Alshanketi, A. Alharbi, M. Kuruvilla, V. Mahzoon, S. T. Siddiqui, N. Rana, and A. Tahir, "Pneumonia detection from chest X-ray images using deep learning and transfer learning for imbalanced datasets," *Journal of Imaging Informatics in Medicine*, vol. 38, pp. 2021–2040, Nov. 2024. https://doi.org/10.1007/s10278-024-01334-0

[9] K. Shahi and A. Bagale, "Weakly supervised pneumonia localization from chest X-rays using deep neural network and grad-cam explanations," *Journal of Artificial Intelligence and Autonomous Intelligence*, vol. 2, no. 3, pp. 450–465, Dec. 2025. https://doi.org/10.54364/JAIAI.2024.1126

[10] E. Yagis, S. W. Atnafu, A. García Seco de Herrera, C. Marzi, R. Scheda, M. Giannelli, C. Tessa, L. Citi, and S. Diciotti, "Effect of data leakage in brain MRI classification using 2D convolutional neural networks," *Scientific Reports*, vol. 11, Nov. 2021, Art. no. 22544. https://doi.org/10.1038/s41598-021-01681-w

[11] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science*, vol 12346, A. Vedaldi, H. Bischof, T. Brox, and J. M. Frahm, Eds. Springer, Cham, 2020, pp. 213–229. https://doi.org/10.1007/978-3-030-58452-8_13

[12] Z. Nabulsi *et al.*, "Deep learning for distinguishing normal versus abnormal chest radiographs and generalization to two unseen diseases tuberculosis and COVID-19," *Scientific Reports*, vol. 11, 2021, Art. no. 15523. https://doi.org/10.1038/s41598-021-93967-2

[13] S. Kim, B. Rim, S. Choi, A. Lee, S. Min, and M. Hong, "Deep learning in multi-class lung diseases classification on chest X-ray images," *Diagnostics*, vol. 12, no. 4, Apr. 2022, Art. no. 915. https://doi.org/10.3390/diagnostics12040915

[14] S. Albahli, H. T. Rauf, A. Algosaibi, and V. E. Balas, "Ai-driven deep CNN approach for multi-label pathology classification using chest X-rays," *PeerJ Computer Science*, vol. 7, Apr. 2021, Art. no. 495. https://doi.org/10.7717/peerj-cs.495

[15] V.-T. Pham, C.-M. Tran, S. Zheng, T.-M. Vu, and S. Nath, "Chest X-ray abnormalities localization via ensemble of deep convolutional neural networks," in *Proceedings of the IEEE International Conference on Advanced Technologies for Communications (ATC)*, Ho Chi Minh City, Vietnam, Oct. 2021, pp. 125–130. https://doi.org/10.1109/ATC52653.2021.9598342

[16] H. T. Nguyen, M. N. Nguyen, S. C. Pham, and P. H. D. Bui, "Abnormalities detection on chest radiograph with bounding box-based lungs extraction and object detection algorithm," *International Journal of Information Technology*, vol. 16, pp. 2241–2251, Feb. 2024. https://doi.org/10.1007/s41870-023-01687-9

[17] N. Ngo, T. Vo, and L. Ngo, "Application of deep learning in chest X-ray abnormality detection," *Vietnam Journal of Science, Technology and Engineering*, vol. 65, no. 4, pp. 84–93, Dec. 2023. https://doi.org/10.31276/VJSTE.65(4).84-93

[18] M. A. Al-antari, C.-H. Hua, J. Bang, and S. Lee, "Fast deep learning computer-aided diagnosis of COVID-19 based on digital chest X-ray images," *Applied Intelligence*, vol. 51, pp. 2890–2907, 2021. https://doi.org/10.1007/s10489-020-02076-6

[19] M. Mustafa and A. Nsour, "Using computer vision techniques to automatically detect abnormalities in chest X-rays," *Diagnostics*, vol. 13, no. 18, Sep. 2023, Art. no. 2979. https://doi.org/10.3390/diagnostics13182979

[20] K. Yu, S. Ghosh, Z. Liu, C. Deible, and K. Batmanghelich, "Anatomy-guided weakly-supervised abnormality localization in chest X-rays," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, *Lecture Notes in Computer Science*, vol. 13435, L. Wang, Q. Dou, P.T. Fletcher, S. Speidel, and S. Li, Eds. Springer, Cham, 2022, pp. 658–668. https://doi.org/10.1007/978-3-031-16443-9_63
[21] T. Nakao *et al.*, "Unsupervised deep anomaly detection in chest radiographs," *Journal of Digital Imaging*, vol. 34, pp. 418–427, Feb. 2021. https://doi.org/10.1007/s10278-020-00413-2
[22] M. Kim, K.-R. Moon, and B.-D. Lee, "Unsupervised anomaly detection for posteroanterior chest X-rays using multiresolution patch-based self-supervised learning," *Scientific Reports*, vol. 13, Feb. 2023, Art. no. 3415, https://doi.org/10.1038/s41598-023-30589-w
[23] H. Sheng, L. Ma, J.-F. Samson, and D. Liu, "BarlowTwins-CXR: enhancing chest X-ray abnormality localization in heterogeneous data with cross-domain self-supervised learning," *BMC Medical Informatics and Decision Making*, vol. 24, May 2024, Art. no. 126. https://doi.org/10.1186/s12911-024-02529-9
[24] C. Hsieh, I. B. Nobre, S. C. Sousa, C. Ouyang, M. Brereton, J. C. Nascimento, J. Jorge, and C. Moreira, "MDF-Net for abnormality detection by fusing X-rays with clinical data," *Scientific Reports*, vol. 13, Sep. 2023, Art. no. 15873. https://doi.org/10.1038/s41598-023-41463-0
[25] N. H. Nguyen, H. Q. Nguyen, N. T. Nguyen, T. V. Nguyen, H. H. Pham, and T. N.-M. Nguyen, "A clinical validation of VinDr-CXR, an AI system for detecting abnormal chest radiographs," *arXiv preprint arXiv:2104.02256*, Apr. 2021. https://doi.org/10.48550/arXiv.2104.02256
[26] F. Behrendt, M. Bengs, D. Bhattacharya, J. Krüger, R. Opfer, and A. Schlaefer, "A systematic approach to deep learning-based nodule detection in chest radiographs," *Scientific Reports*, vol. 13, June 2023, Art. no. 10120. https://doi.org/10.1038/s41598-023-37270-2
[27] Y. Xie, B. Zhu, Y. Jiang, B. Zhao, and H. Yu, "Diagnosis of pneumonia from chest X-ray images using YOLO deep learning," *Frontiers in Neurorobotics*, vol. 19, Apr. 2025, Art. no. 1576438. https://doi.org/10.3389/fnbot.2025.1576438
[28] H. Q. Nguyen *et al.*, "VinBigData chest X-ray abnormalities detection," Kaggle, 2020. [Online]. Available: https://kaggle.com/competitions/vinbigdata-chest-xray-abnormalities-detection
[29] L. Sun, C. Xia, W. Yin, T. Liang, P. S. Yu, and L. He, "Mixup-transformer: Dynamic data augmentation for NLP tasks," in *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, Spain, 2020, pp. 3436–3440. https://doi.org/10.18653/v1/2020.coling-main.305
[30] X. Xu, B. Zhao, X. Tong, H. Xie, Y. Feng, C. Wang, C. Xiao, X. Ke, and J. Du, "A data augmentation strategy combining a modified pix2pix model and the copy-paste operator for solid waste detection with remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 8484–8491, Sep. 2022. https://doi.org/10.1109/JSTARS.2022.3209967
[31] M. Yurdakul, H. Sazak, M. Kotan, and Ş. Taşdemir, "A review of YOLO family from YOLOv1 to YOLO26," *Preprints.org*, Feb. 2026. https://doi.org/10.20944/preprints202602.1844.v1
[32] K. Jiang, T. Xie, R. Yan, X. Wen, D. Li, H. Jiang, N. Jiang, L. Feng, X. Duan, and J. Wang, "An attention mechanism-improved YOLOv7 object detection algorithm for hemp duck count estimation," *Agriculture*, vol. 12, no. 10, Oct. 2022, Art. no. 1659. https://doi.org/10.3390/agriculture12101659
[33] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable transformers for end-to-end object detection," *arXiv preprint arXiv:2010.04159*, Oct. 2020. https://doi.org/10.48550/arXiv.2010.04159
[34] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, Niteroi, Brazil, Yuly 2020, pp. 237–242. https://doi.org/10.1109/IWSSIP48289.2020.9145130
[35] L. Zhao and S. Li, "Object detection algorithm based on improved YOLOv3," *Electronics*, vol. 9, no. 3, Mar. 2020, Art. no. 537. https://doi.org/10.3390/electronics9030537
[36] S. Tang, S. Zhang, and Y. Fang, "HIC-YOLOv5: Improved YOLOv5 for small object detection," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, May 2024, pp. 6614–6619. https://doi.org/10.1109/ICRA57147.2024.10610273
[37] E. Liang, D. Wei, F. Li, H. Lv, and S. Li, "Object detection model of vehicle-road cooperative autonomous driving based on improved YOLO11 algorithm," *Scientific Reports*, vol. 15, Sep. 2025, Art. no. 32348. https://doi.org/10.1038/s41598-025-18263-9
[38] C.-L. Fan, "Evaluation model for crack detection with deep learning: Improved confusion matrix based on linear features," *Journal of Construction Engineering and Management*, vol. 151, no. 3, Mar. 2025. https://doi.org/10.1061/JCEMD4.COENG-14976
[39] S. Wu, J. Yang, X. Wang, and X. Li, "IoU-balanced loss functions for single-stage object detection," *Pattern Recognition Letters*, vol. 156, pp. 96–103, Apr. 2022. https://doi.org/10.1016/j.patrec.2022.01.021

Thi-Da-Huong Truong is an undergraduate student at the College of Information and Communication Technology, Can Tho University, Vietnam (2022–2026), majoring in Information Systems. Her academic interests include data science, big data technologies, computer vision, machine learning, deep learning, data-driven decision-making systems, and smart healthcare systems.
Email: huongtruong290104@gmail.com

Ngoc Huynh Pham is a Teaching Assistant at the College of Information and Communication Technology, Can Tho University, Vietnam. She received her Master degree in 2025. She has a deep passion and interest in information technology, especially in areas such as applications based on machine learning and computer vision.
Email: phngoc@ctu.edu.vn

Vo-Phuong-Tam Nguyen is an undergraduate student at Can Tho University, Vietnam, from 2022 to 2026. She is majoring in Information Systems. Her academic interests include database management, systems analysis and design, and digital transformation. She is particularly interested in applying machine learning and data analytics to optimise business processes.
Email: nguyenvophuongtam@gmail.com

Thi-Thanh-Thuy Le is an undergraduate student at Can Tho University, Vietnam (2022–2026), majoring in Information Systems. Her academic interests include data analytics, big data technologies, and machine learning. She is particularly interested in exploring data analysis techniques and machine learning methods for analysing large-scale datasets.
Email: thanh231206thuy@gmail.com

Hai Thanh Nguyen is an Associate Professor at the College of Information and Communication Technology, Can Tho University, Vietnam. He received his Master's degree in Computer Science and Engineering from National Chiao Tung University, Taiwan, and his PhD degree in Computer Science from Sorbonne University, France. His current research includes healthcare systems and applications using computer vision, machine learning, and data analysis.
Email: nthai.cit@ctu.edu.vn
ORCID iD: https://orcid.org/0000-0002-1386-1390