Research Paper

Context-Aware Hybrid Attention-Enhanced YOLO Framework for Accurate Multi-Scale Scoliosis Detection Using X-Ray Images

Duraikannu Murugasen Kalai Selvi*, Periyathambi Ezhumalai

Department of Computer Science and Engineering, R.M. D Engineering College, R.S.M Nagar, Kavaraipettai, Gummidipoondi Taluk, Tiruvallur District, Tamil Nadu, India

Citation Selvi DMK, Ezhumalai P. Context-Aware Hybrid Attention-Enhanced YOLO Framework for Accurate Multi-Scale Scoliosis Detection Using X-Ray Images. International Journal of Medical Toxicology and Forensic Medicine. 2026; 16:E52666.

https://doi.org/10.22037/ijmtfm.v16.52666

Article info:

Received: 01 July, 2026

Accepted: 15 July, 2026

Published: 21 August, 2026

Keywords:

Scoliosis detection, Deep learning, X-ray images, Spinal curvature, AdamW optimizer, Explainable artificial intelligence

ABSTRACT

Background: Scoliosis is a condition in which the spine curves abnormally, and it can be difficult to identify early because the changes are often subtle and the spine's structure is complex. Traditional manual diagnosis is time-consuming and depends on clinical expertise. Therefore, this study presents a Hybrid Attention-Enhanced YOLO Framework for Accurate Multi-Scale Scoliosis Detection (HAYOLO-MSD) approach.

Methods: The proposed model uses a hybrid CNN–Transformer backbone with C2f blocks to capture both local details and global context. A feature fusion module combines these representations for better learning. In the neck, an attention-based APAN replaces traditional FPN and PAN to focus on spinal curvature and abnormalities. The decoupled head separates classification and regression tasks, with attention-enhanced Softmax for prediction. Training is optimized using AdamW, and Grad-CAM provides interpretability by highlighting important regions.

Results: Extensive experiments on the Balanced Scoliosis X-ray Dataset demonstrate strong performance of the proposed system. The comparative result analysis demonstrates that the HAYOLO-MSD method achieves 96.56% accuracy on the Balanced Scoliosis X-ray Dataset and 95.73 accuracy on the Mendeley dataset.

Conclusion: The proposed HAYOLO-MSD approach provides an effective and accurate framework for multi-scale scoliosis detection, demonstrating strong performance on both the Balanced Scoliosis X-ray dataset and the Mendeley dataset.

*** Corresponding Author:**

Duraikannu Murugasen Kalai Selvi

Department of Computer Science and Engineering, R.M. D Engineering College, R.S.M Nagar, Kavaraipettai, Gummidipoondi Taluk, Tiruvallur District, Tamil Nadu, India.

E-mail: *dmkalai@gmail.com*

Introduction

Scoliosis is an irregular curvature of the spine that commonly occurs in teenagers. Because of the quick spinal growth during this stage, an increase in scoliosis might often be recognized by radiographs of the spine in the sagittal planes and posteroanterior. Treatment decisions are determined based on the type and severity of the condition. Delayed treatment can damage lung growth, resulting in pulmonary hypertension. The method of identifying spinal column disease starts with a manual assessment. Once an early diagnosis is made, the next step would be medical diagnostic imaging. X-ray images of the back deliver additional data about the anatomy of the spine and the presence of fractures, infections, or other irregularities [1]. Though conventional radiologic models like X-rays expose patients to higher dosages of radiation, particularly adolescents and children, posing long-term health issues. To overcome these limitations, the fast growth of artificial intelligence (AI) techniques with deep learning (DL) and machine learning (ML) approaches transforming medical image analysis because of their capacity to process large datasets with higher precision [2].

A brief related works on scoliosis detection. In, an innovative auxiliary diagnostic technique for scoliosis was formulated on hierarchical contrastive learning [3]. The authors developed the ScoliGaitX, a new non-invasive model for evaluating scoliosis by employing video analysis of gait [4]. Duan et al. [5] implemented a DL-based image segmentation technique to improve the effectiveness of scoliosis identification. Mulford et al. [6] presented the DL based image classification for greater performance in radiographic categorization of scoliosis etiology when related to qualified spine surgeons reliant on spine radiograms. Natesan et al. [7] concentrated on the identification and classification of spine irregularities by employing the innovative DL techniques. In introduced a strong and XAI model by employing DL techniques for predicting the development of scoliosis via only vertical frontal radiographs [8]. Feng et al. [9] addressed the restrictions of radiographic imaging methods and single-task learning techniques by emerging the non-invasive, radiation-free diagnostic technique. Maharasi et al. [10] implemented an innovative DL method for identification of vertebrae landmarks and scoliosis diagnosis. It is the spinal abnormalities that can be represented through the distortion of the spine structures.

Existing DL techniques are effective for scoliosis detection, but still struggle with background noise, small abnormal region detection, and efficient feature fusion, while also failing to balance accuracy and computational cost. The following are key contribution of this research:

- Design a YOLOV8-based convolutional neural network (CNN) as the backbone, combined with multiple C2f blocks and a Transformer model to effectively capture both local spatial features and global contextual spinal information.

- Develop feature fusion strategy integrates CNN and Transformer outputs to combine fine-grained details with high-level semantic features.

- Introduce an attention-based path aggregation network (APAN) and a decoupled detection head with convolutional block attention mechanism (CBAM) are introduced to enhance multi-scale representation and improve localization.

Employ an efficient loss function with AdamW optimization for stable training and Grad-CAM for interpretable visualization of key regions influencing detection and classification performance.

Materials and Methods

Cell Line and Culture Conditions

This research presents a systematic pipeline for scoliosis classification leveraging DL methodologies. The overall workflow of the HAYOLO-MSD approach is depicted in Figure 1. By using a YOLO-based CNN with C2f blocks and a Transformer. Feature fusion and APAN improve multi-scale representation and reduce noise. A decoupled head handles detection and classification, while AdamW ensures stable training. Grad-CAM is used for visual interpretability.

Dataset Used

The proposed technique is examined by employing the Balanced Scoliosis X-ray dataset [11]. This is an augmented scoliosis X-ray dataset with four classes used for deep learning tasks Table 1 shows that the dataset consists of a total of 5,963 samples with classes.

CA-YOLO++: Context-Aware Hybrid YOLO with Decoupled Attention Head

The CA-YOLO++ uses a CNN–Transformer backbone for global feature learning. APAN and feature fusion improve multi-scale representation and focus on spinal curvature [13,14]. A decoupled head enhances accuracy by separating regression and classification.

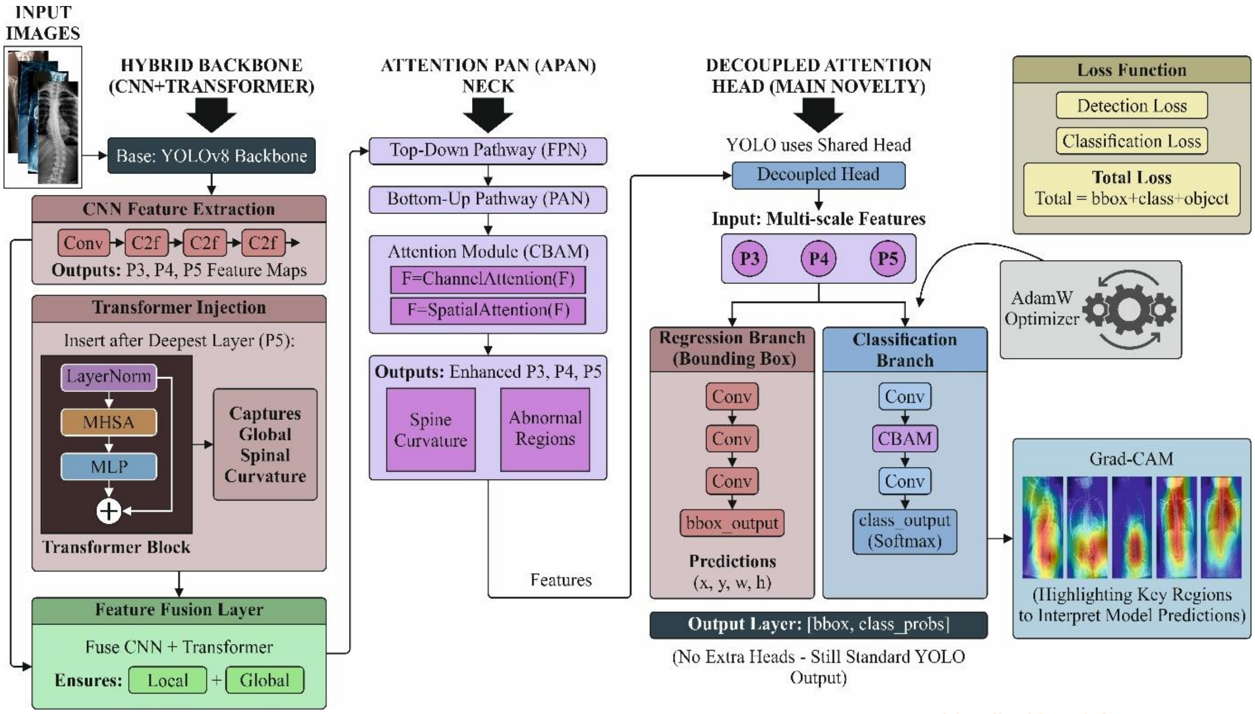

Figure 1. Overall workflow of the propose method.

Table 1. Details of the dataset.

Classes	Sample Images
Mild Scoliosis	1439
Moderate Scoliosis	1582
Severe Scoliosis	1523
Healthy	1419
Total	**5963**

Hybrid Backbone

The model employs a YOLOv8 backbone as its base for feature representation. It combines CNN layers for local feature learning and a Transformer block for extracting global context. YOLOv8 is a modern object detection model that balances accuracy and speed. It processes input images through a backbone network to extract multi-scale visual features, which are then refined and fused for better representation. Its modular design makes it efficient and suitable for real-time detection tasks. Transformer structure involves Multi-Head Self-Attention (MHSA), Multi-Layer Perceptron (MLP), residual connections, and LayerNorm. MHSA uses multiple parallel heads, where each processes Query (Q), Key (K), and Value (V) through independent linear projections. LayerNorm are computed for the attention scores across all heads and then ordered. The attention head with the greatest norm value has been identified having the maximum attention chances. MLP predicts the Remaining Useful Life (RUL) by taking degradation features as input.

Attention Neck

The APAN integrates PAN with attention mechanisms to effectively aggregate multi-scale features. PAN is utilized to improve feature fusion by integrating bottom-up and top-down (FPN) and pathways, facilitating better propagation of lower and higher-level characteristics.

Attention module enhances feature maps by sequentially using channel and spatial attention, enabling better focus on significant areas. It includes the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). The pooled features are passed through a shared MLP and Sigmoid activation to generate the channel attention map M_c. The feature map is refined as follows:

$$M_c(F) = \sigma\Big(\mathrm{MLP}\big(\mathrm{AvgPool}(F)\big) + \mathrm{MLP}\big(\mathrm{MaxPool}(F)\big)\Big)\,\sigma\Big(W_1\big(W_0(F^c_{\mathrm{avg}})\big) + W_1\big(W_0(F^c_{\mathrm{max}})\big)\Big) \quad (1)$$

$$F' = M_c(F) \otimes F \quad (2)$$

SAM applies average and max pooling followed by a 7×7convolution and Sigmoid activation to generate M_s, which is used to refine F'into the final feature map.

$$M_s(F) = \sigma\left(f^{7\times7}([\text{AvgPool}(F); \text{MaxPool}(F)])\right) = \sigma\left(f^{7\times7}\left([F_{\text{avg}}^{S}, F_{\text{max}}^{S}]\right)\right) \quad (3)$$

$$F'' = M_s(F') \otimes F' \quad (4)$$

The decoupled attention head uses separate branches for regression and classification to improve task-specific learning. The Classification Branch uses a Softmax layer to convert features into class probabilities of scoliosis severity or categories. The final output is obtained using Softmax to generate class probabilities for each target region. The loss function employs CIoU Loss for detection and Focal Loss with Label Smoothing for classification, combined as total loss.

Detection Loss: The total loss contains 3 elements:

Bounding Box Regression Loss:

$$L_{\text{box}} = \text{CIoU Loss} \quad (5)$$

Classification Loss:

$$L_{\text{cls}} = \text{Focal Loss} + \text{Label Smoothing} \quad (6)$$

Total Loss:

$$L_{\text{total}} = L_{\text{box}} + L_{\text{cls}} + L_{\text{obj}} \quad (7)$$

Here, L_{obj} refers to the objectness loss, and the integration of entire 3 elements.

Parameter Optimization

To further enhance the classification method, the AdamW optimizer enhances training stability and speeds up convergence while improving generalization of the model [15]. AdamW applies weight decay minimizes the weights directly in the upgrade stage rather than varying the loss function, it instantly in parameter upgrades that improves both convergence and generalization:

$$w_{t+1} = w_t - \eta(\text{AdamGradient}) - \eta\lambda w_t \quad (8)$$

Grad-CAM-based Model Interpretability

Grad-CAM is a visualization model that highlights important regions in an image to interpret model decisions [16]. Considered an input image $I \in \mathbb{R}^{3\times H\times W}$ and a network: $I \mapsto y$, here $y \in \mathbb{R}^{C}$ represents the class scores for Ccategories, y^cdenotes the pre-softmax score for class c, and $A_k \in \mathbb{R}^{M\times N}$is the activation map of the k^{th}convolutional channel. The importance weights are obtained using gradients of $y^c$with respect to feature maps and global average pooling:

$$S_k = \left(\frac{1}{Z}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\frac{\partial y^c}{\partial A_k}\right)_{ij}\right) \odot A_k \quad (9)$$

Here, $\frac{\partial y^c}{\partial A_k}$ represents the gradient map, and $Z$is a normalization factor. The final Grad-CAM heatmap is computed as:

$$L_{GradCAM} = ReLU\left(\sum_{k=1}^{K} S_k\right) \quad (10)$$

This produces a saliency map highlighting region that positively contribute to the target class prediction.

Results

The performance assessment of proposed model is inspected using Balanced Scoliosis X-ray Dataset and Mendeley datatset. Figure 2 depicts the sample of original and annotated images. Figure 3 displays predicted X-ray images of scoliosis

Figure 4 displays the training (TRAING) and validation (VALIDN) loss trends for box loss, classification loss, and distribution focal loss over 50 epochs. The consistent decrease in both TRAING and VALIDN losses indicates effective learning and good generalization with minimal overfitting.

Table 2 present the object detection result of the HAYOLO-MSD approach with other techniques. The proposed method has achieved superior performance with $preci_n$ of 89.14%, $recal_l$ of 93.82%, mAP@50 of 93.11%, and mAP@50-95 of 81.46%.

Figure 5 illustrate the training metrics of the proposed model, including key detection metrics like precision, recall, mAP@50, and mAP@50–9 over 50 training epochs. All metrics demonstrate a constant upward trend, with recall increasing the faster, and then precision and mAP scores.

Figure 6 shows the classifier outcomes of the HAYOLO-MSD method. Fig. 6a exhibit the confusion matrix with precise recognition and classification of four class labels. Finally, Fig. 6b shows the ROC analysis, signifying proficient outcomes with higher ROC values for diverse class labels.

Figure 7 represent the classifier outcomes of the HAYOLO-MSD technique at 80:20. The experimental outcome values are indicated that the proposed system gains average value with $accur_y$ of 95.57% and 95.39% for the TRAP and TESP.

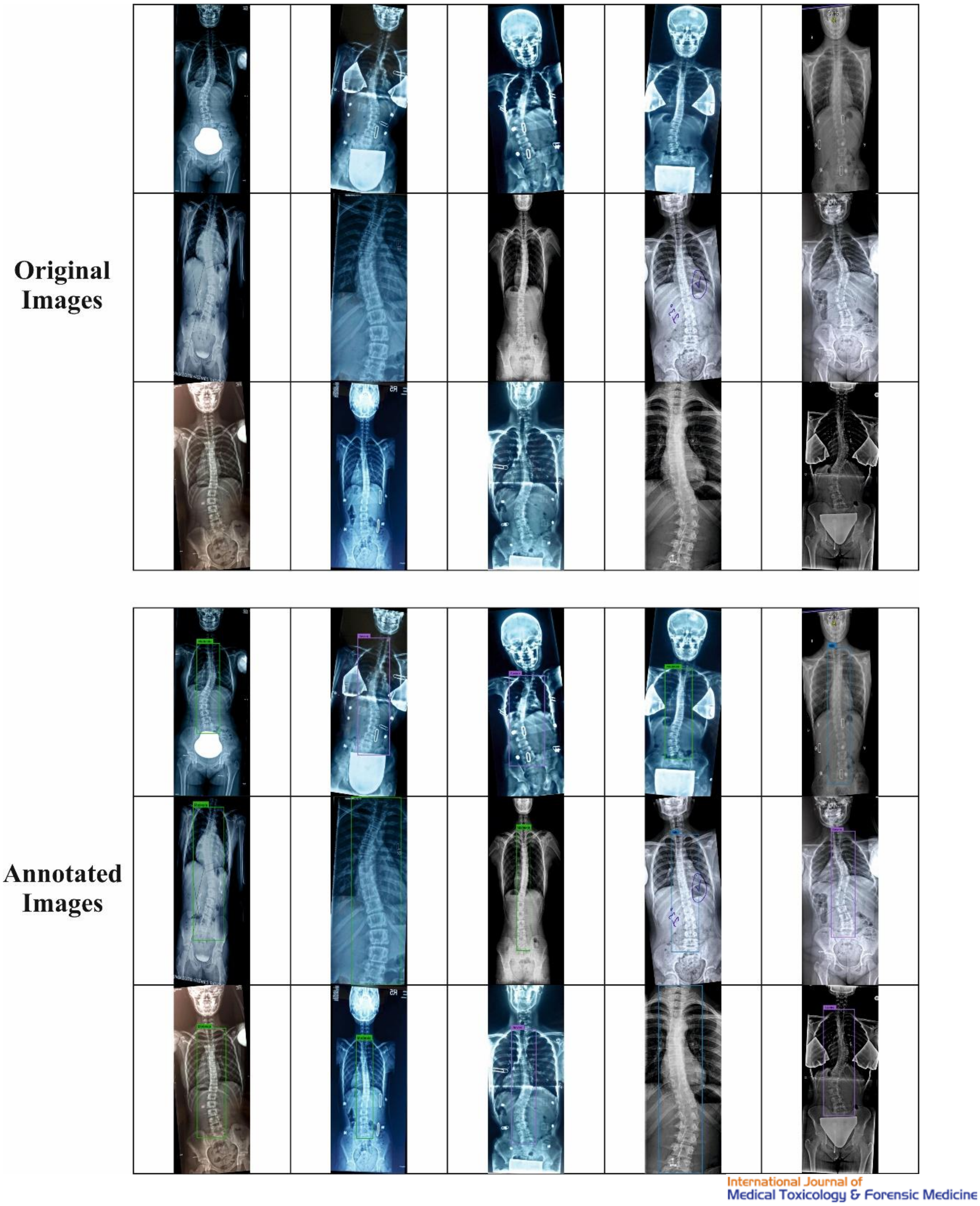

Figure 2. Sample of original and annotated images.

Discussion

The comparison evaluation of proposed approaches with existing methods are represented in Figure 8 [17-18]. The outcome implied that the GoogLeNet, VGG-16, ResNet and Faster R-CNN ResNet-101, EfficientNet, Auto Spine-Net, the ROC and Cobb approach has achieved lower value, while the proposed HAYOLO-MSD method has reached maximal value with $accur_y$ of 96.56%, $preci_n$ of 93.19%, $recal_l$ of 92.99%, $F1_{score}$ of 93.06%.

The comparison evaluation of proposed approaches with existing methods are represented in Table 3 [19-20]. The outcome implied that the Efficient-CapsNet, DenseNet-201, SqueezeNet, Mask RCNN and YOLOv5, and AlexNet approach has achieved lower value, while the proposed HAYOLO-MSD method has reached maximal value with Accuracy of 95.73%, Precision of 92.42%, Recall of 92.04%, F1-Score of 92.23 and execution time of 0.68 sec.

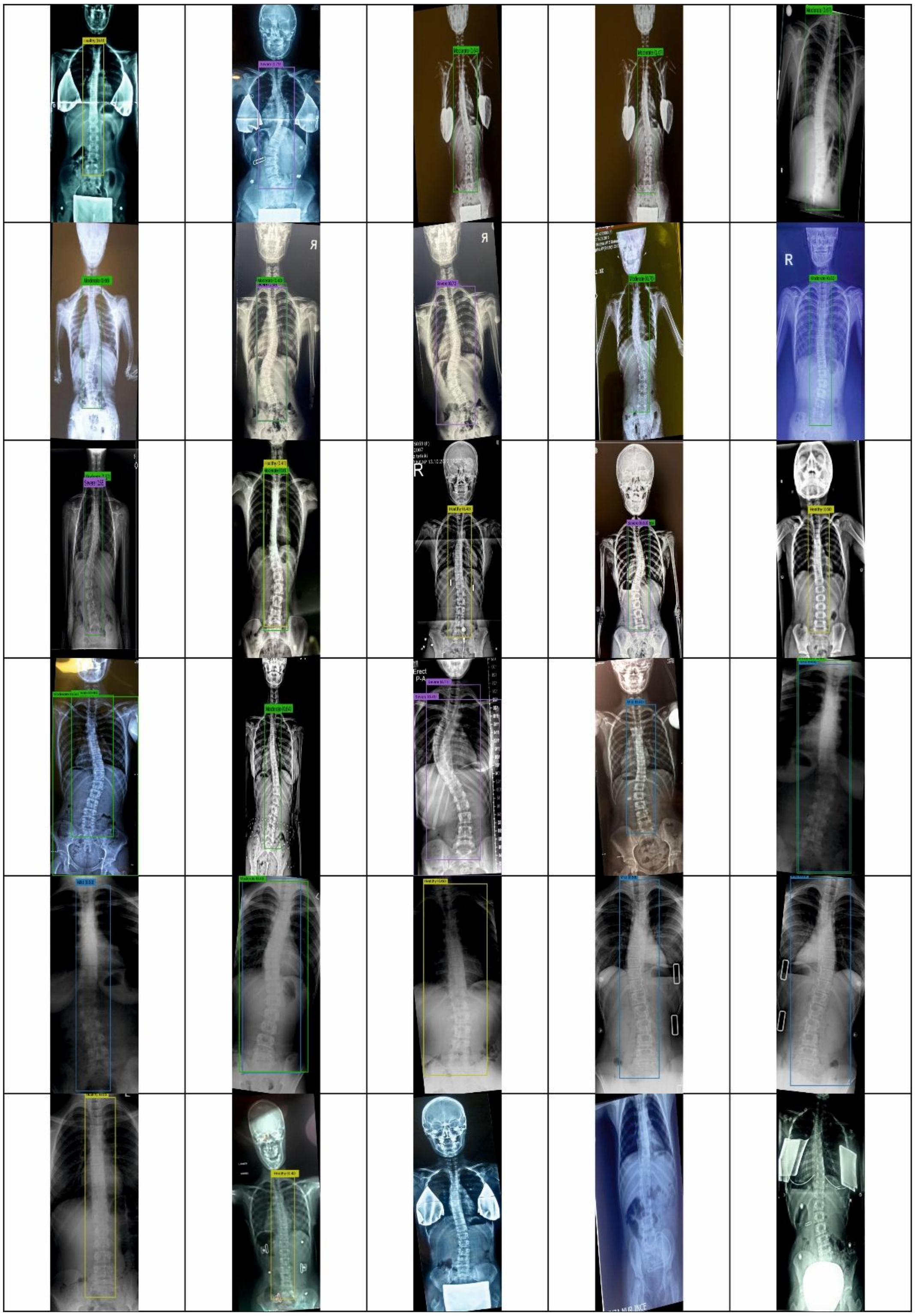

Figure 3. Predicted X-ray Images of Scoliosis.

Figure 9 presents attention-based heatmaps over spinal X-ray images, where warmer regions indicate higher model focus. These visualizations exhibit how the model captures curvature patterns and asymmetries along diverse parts of the spine. The highlighted areas correspond to clinically relevant regions used for scoliosis assessment and diagnosis.

Conclusion

This study has projected a HAYOLO-MSD approach for identifying and classifying scoliosis from X-ray images. It combines a CNN–Transformer

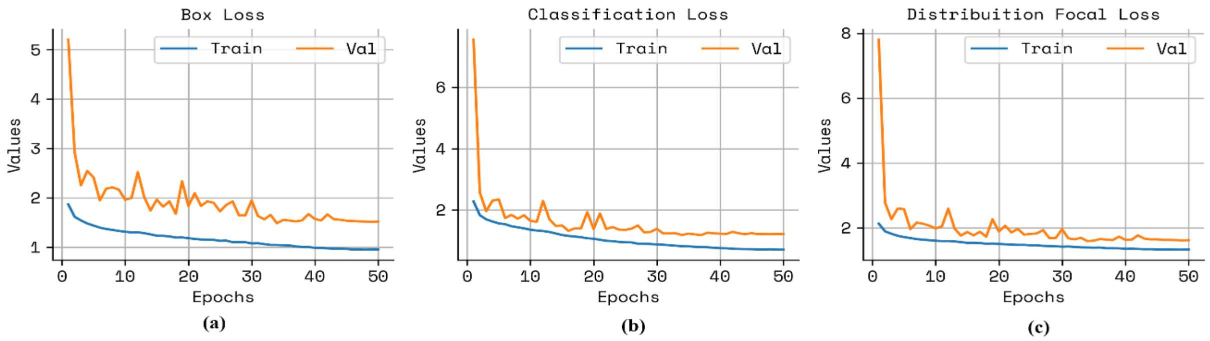

Figure 4. Loss curve with (a) Box, (b) Classification, and (c) Distribution focal.

Table 2. Object detection of the proposed method.

Models	$Preci_n$	$Recal_l$	mAP@50	mAP@50-95
YOLOv5s	87.32	54.54	64.93	63.37
YOLOv5m	45.78	92.12	85.70	77.26
YOLOv5I	73.58	86.90	91.77	59.33
YOLOv5x	79.79	57.73	78.59	79.75
YOLOv8n-OBB	83.90	72.62	74.78	60.75
C2f-DCNv2+OBB	87.50	47.52	68.09	51.91
YOLOv8-DSF	47.88	82.80	52.02	69.69
Proposed	89.14	93.82	93.11	81.46

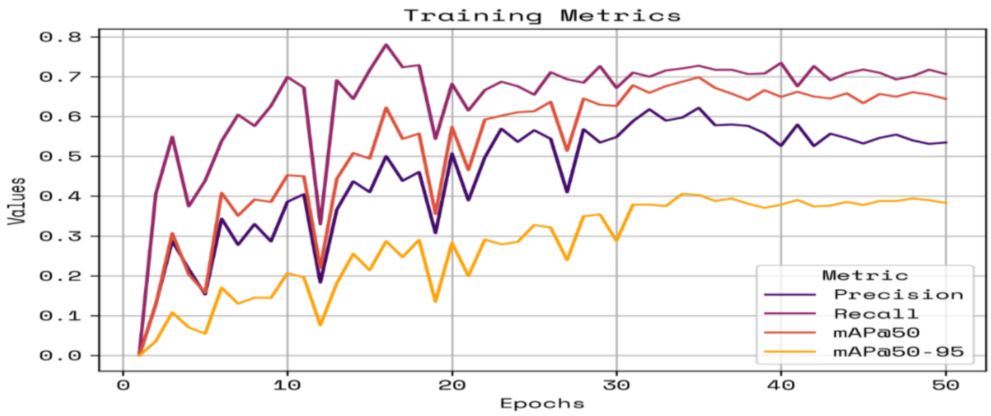

Figure 5. Training metrics of the HAYOLO-MSD model.

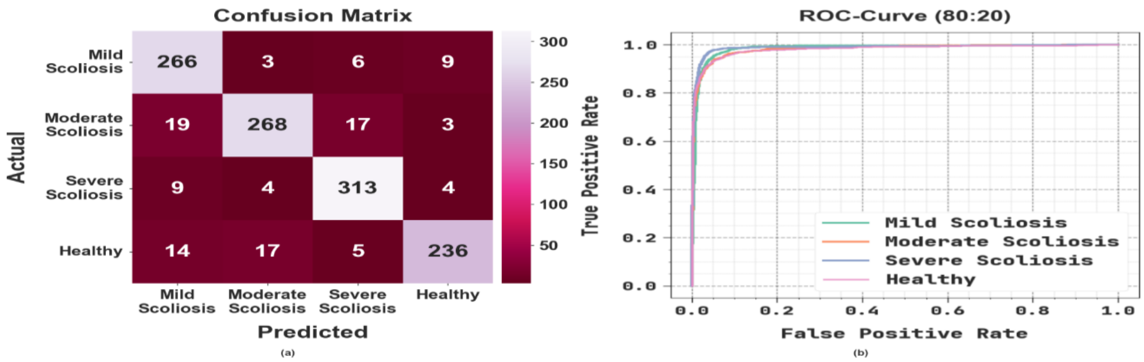

Figure 6. Classifier result with (a) Confusion matrix and (b) ROC-curve.

backbone with feature fusion and an attention-based APAN to better capture spinal abnormalities. A decoupled head improves classification and localization, while AdamW ensures stable training, and

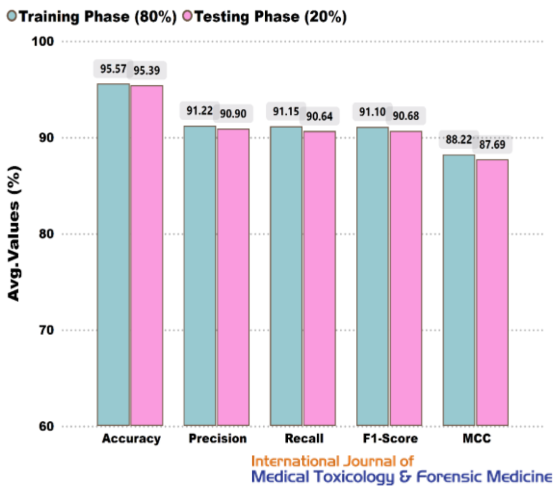

Figure 7. Average value of the proposed method with 80:20.

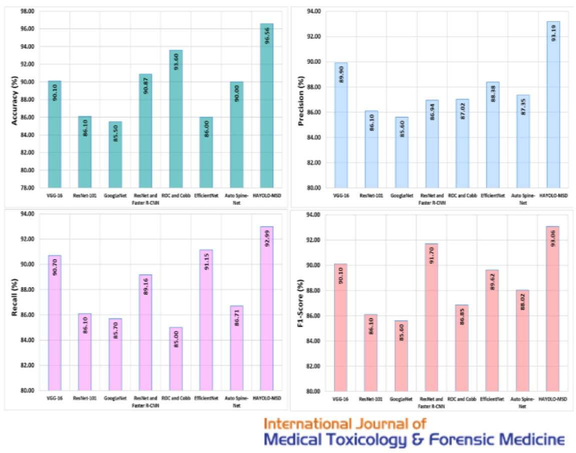

Figure 8. Comparative outcomes of the HAYOLO-MSD method with measure.

Table 3. Comparative outcomes of the HAYOLO-MSD method with measure.

Methodologies	Accuracy	Precision	Recall	F1-Score	Execution Time (sec)
Efficient-CapsNet	77.10	74.83	72.46	73.63	1.84
DenseNet-201	94.31	90.76	89.48	90.12	1.57
SqueezeNet	93.64	89.53	88.76	89.14	0.93
Mask RCNN and YOLOv5	94.69	91.38	90.72	91.05	1.42
AlexNet	91.67	87.24	85.69	86.46	1.18
HAYOLO-MSD	**95.73**	**92.42**	**92.04**	**92.23**	**0.68**

International Journal of Medical Toxicology & Forensic Medicine

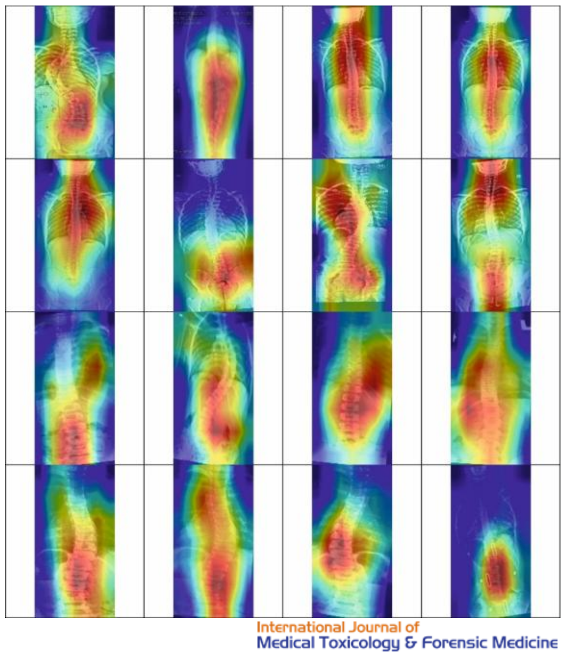

Figure 9. Grad-CAM heatmap visualization with various samples.

Grad-CAM adds interpretability. The model achieves 96.56% accuracy on the Balanced Scoliosis X-ray Dataset and 95.73 accuracy on the Mendeley dataset, outperforming existing methods. Despite improved accuracy, the proposed model has limitations, and future work will focus on lightweight, energy-efficient models and adaptive ensembles to enhance robustness and real-world medical applicability.

In the future, lightweight model compression techniques, such as knowledge distillation, quantization, and pruning, will be explored to reduce computational complexity and memory requirements. This will ease the efficient deployment of the proposed HAYOLO-MSD architecture on resource-constrained clinical devices.

Ethical Approval

None.

Acknowledgment

None.

Funding

No external funding was received for this research.

Conflicts of Interest

The authors report there are no competing interests to declare.

References

[1] Martrenchar P, Caille A, Dina A, Odent T. A preliminary report: screening idiopathic scoliosis (IS) with a smartphone application. J Pediatr Orthop. 2026:10-1097. [DOI: 10.1097/BPO.0000000000003228]

[2] Yeşilmen N, Baydoğan MP, Danacı Ç, Tuncer SA, Çınar A, Tuncer T. Multi-class idiopathic scoliosis detection based on Cobb angle using a hybrid ViT model. Signal Image Video Process. 2025;19(14):1225. [DOI: 10.1007/s11760-025-04810-4]

[3] Li R, Ye B, Wang W, Liu P, Xiao B, Liu Z, et al. Hierarchical contrastive learning framework for digital scoliosis screening standards, efficacy evaluation and real-world deployment. SSRN Electron J. 2026 Jan. [DOI: 10.2139/ssrn.6210365]

[4] Vishwakarma K, Nigam A. ScoliGaitX: a deep multi-modal fusion network for scoliosis assessment via gait video analysis. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision; 2026. p. 2404-13. [DOI: 10.1109/WACV61042.2026.00237]

[5] Duan X, Ma X, Zhu M, Wang L, You D, Deng L, et al. Deep learning-assisted screening and diagnosis of scoliosis: segmentation of bare-back images via an attention-enhanced convolutional neural network. J Orthop Surg Res. 2025;20(1):161. [DOI: 10.1186/s13018-025-05564-y]

[6] Mulford KL, Nugraha HK, Todderud JE, Pumford A, Milbrandt TA, Huddleston PM, et al. Predicting nonidiopathic scoliosis from plain radiographs: a deep-learning approach. JBJS Open Access. 2025;10(4):e25. [DOI: 10.2106/JBJS.OA.25.00257]

[7] Natesan G, Murugesan A, Nagarajan R. Lightweight CNN-enabled framework for automated detection of scoliosis and spondylosis using spine X-rays. J Comput Cogn Eng. 2025. [DOI: 10.47852/bonviewJCCE52026933]

[8] Takahashi S, Ichikawa S, Watanabe K, Ueda H, Arima H, Yamato Y, et al. Development and validation of transformer- and convolutional neural network-based deep learning models to predict curve progression in adolescent idiopathic scoliosis. J Clin Med. 2025;14(20):7216. [DOI: 10.3390/jcm14207216]

[9] Feng CK, Chen YJ, Dinh QT, Tran KT, Liu CY. Multi-task learning for classification and prediction of adolescent idiopathic scoliosis based on fringe-projection three-dimensional imaging. Eur Spine J. 2025;34(11):5321-33. [DOI: 10.1007/s00586-025-09340-8]

[10] Maharasi M, Senthilnayaki N, Snehaprabha K. Vertebrae landmark detection and scoliosis assessment using deep learning. In: 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT); 2024. p. 1-6. [DOI: 10.1109/IC3IoT60841.2024.10550261]

[11] Salmankey. Balanced scoliosis X-ray dataset, YOLOv5 format [Internet]. Kaggle. Available from: https://www.kaggle.com/datasets/salmankey/balanced-scoliosis-x-ray-dataset-yolov5-format/

[12] Mendeley Data. Dataset xkt857dsxk/1 [Internet]. Available from: https://data.mendeley.com/datasets/xkt857dsxk/1

[13] Ma S, Lu H, Liu J, Zhu Y, Sang P. Layn: lightweight multi-scale attention YOLOv8 network for small object detection. IEEE Access. 2024;12:29294-307. [DOI: 10.1109/ACCESS.2024.3368848]

[14] Farag MM, Fouad M, Abdel-Hamid AT. Automatic severity classification of diabetic retinopathy based on DenseNet and convolutional block attention module. IEEE Access. 2022;10:38299-308. [DOI: 10.1109/ACCESS.2022.3165193]

[15] Zheng H, Wang B, Xiao M, Qin H, Wu Z, Tan L. Adaptive friction in deep learning: enhancing optimizers with sigmoid and tanh function. In: 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS); 2024. p. 809-13. [DOI: 10.1109/ICPICS62053.2024.10797007]

[16] Yang Y, Guo R, Wu S, Wang Y, Yang L, Fan B, et al. DecomCAM: advancing beyond saliency maps through decomposition and integration. Neurocomputing. 2024;610:127826. [DOI: 10.1016/j.neucom.2024.127826]

[17] Zhang L, Pei B, Zhang S, Lu D, Xu Y, Huang X, et al. A new method for scoliosis screening incorporating deep learning with back images.

Global Spine J. 2025;15(4):2062-74. [DOI: 10.1177/21925682241282581]

[18] Lu QT, Nguyen TM. An approach to classifying X-ray images of scoliosis and spondylolisthesis based on fine-tuned Xception model. Int J Adv Comput Sci Appl. 2024;15(2). [DOI: 10.14569/IJACSA.2024.0150262]

[19] Lu QT, Nguyen TM. An approach to classifying X-ray images of scoliosis and spondylolisthesis based on fine-tuned Xception model. Int J Adv Comput Sci Appl. 2024;15(2). [DOI: 10.14569/IJACSA.2024.0150262]

[20] Yeşilmen N, Danacı Ç, Baydoğan MP, Tuncer SA, Çınar A, Tuncer T. Enhanced vision transformer with custom attention mechanism for automated idiopathic scoliosis classification. J Imaging Inform Med. 2026;39(2):1122-39. [DOI: 10.1007/s10278-025-01564-w]