Automated assessment of cervical vertebral maturation stages on lateral cephalometric radiographs using a two-stage deep learning framework

Maryam Javaheri Mahd
Shahid Sadoughi University of Medical Sciences
Hossein Agha Aghili
Shahid Sadoughi University of Medical Sciences
Hamid Reza Dehghan
Shahid Sadoughi University of Medical Sciences
Ali Kamaei
Shahid Sadoughi University of Medical Sciences
Aref Mohammadi Fard
Shahid Sadoughi University of Medical Sciences
Yaser Safı
Shahid Beheshti University of Medical Sciences
Mahjube Entezar-e-Ghaem
Shahid Sadoughi University of Medical Sciences
Mohadese Zare Ebrahimabad
mozr2010@gmail.com

Shahid Sadoughi University of Medical Sciences

Article

Keywords: Cervical Vertebral Maturation (CVM), Deep Learning, Mask R-CNN, EfficientNet, Cephalometric

Posted Date: May 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9226756/v1

Additional Declarations: No competing interests reported.

Abstract

Background

Accurate assessment of cervical vertebral maturation (CVM) is critical for timing orthodontic interventions. Manual assessment is often subjective and prone to inter-observer variability. This study aimed to develop and validate a fully automated, two-stage deep learning framework for CVM stage classification using lateral cephalometric radiographs (LCRs).

Methods

A dataset of 1102 LCRs from individuals aged 7–18 years was curated, preserving native radiographic noise to accurately simulate real-world imaging conditions. The proposed pipeline consists of two stages: (1) automated detection and instance segmentation of C2, C3, and C4 vertebrae using Mask R-CNN architecture with a ResNet-50-FPN backbone, and (2) skeletal maturity classification (CS1–CS6) using an EfficientNet-B3 model with a late-fusion strategy. The model was trained using transfer learning and evaluated using mean average precision (mAP), accuracy, and confusion matrices.

Results

The detection model achieved high localization precision (AP@IoU = 0.5 = 0.96; mAP@50 = 0.85). The classification stage demonstrated an overall accuracy of approximately 70%, with peak performance in identifying CS6 (73%). While early (CS1-CS2) and late stages showed high reliability, the transitional stage (CS3) exhibited the lowest accuracy (33%), reflecting the inherent morphological overlap during peak pubertal growth.

Conclusion

The proposed framework provided a standardized, reproducible, and fully automated tool for skeletal maturity assessment. Its robustness against real-world image noise and high anatomical accuracy can make it a suitable auxiliary tool for orthodontic clinical diagnoses.

Introduction

Accurate assessment of skeletal maturation is a fundamental component of clinical decision-making across dental disciplines, most notably orthodontics, maxillofacial surgery, and implantology. While chronological age is easily determined, skeletal maturation stages offer a far more reliable correlation with the adolescent growth spurt. This correlation allows clinicians to pinpoint the optimal window for treatment and more accurately predict outcomes that depend on growth [1,2]. For pediatric and adolescent patients, timing orthodontic or orthopaedic interventions to coincide with the pubertal growth peak can significantly enhance treatment efficacy. Conversely, poorly timed treatment often leads to relapse, unnecessary biological costs, and extended treatment durations. Consequently, precise determination of skeletal maturity is not merely an adjunct but a prerequisite for high-quality, growth-oriented treatment

planning [3,4]. Historically, skeletal maturation has been evaluated through various methods, with hand–wrist radiography and cervical vertebral maturation (CVM) analysis via lateral cephalometric imaging being the most prominent [5]. Over the last two decades, CVM analysis has largely surpassed hand-wrist methods in clinical popularity because it utilises routine orthodontic records, thereby sparing the patient from additional radiation exposure. Despite its utility, the method is inherently limited by its reliance on subjective visual interpretation of vertebral morphology. Clinicians must assess subtle changes in the concavity of the lower borders and the shape of the C2, C3, and C4 vertebrae. This subjectivity often results in significant inter-examiner variability and inconsistent reproducibility, which can undermine diagnostic reliability [6,7]. To address these human limitations, researchers have long sought to automate the CVM staging process. Early attempts utilised traditional computer vision techniques, such as Active Shape Models (ASMs) and manual feature extraction, to identify vertebral landmarks. While these were pioneering, they often struggled with the high levels of noise and varying contrast typical of lateral cephalograms. With the recent breakthrough of deep learning, specifically Convolutional Neural Networks (CNNs), the landscape of automated diagnostics has shifted. Unlike traditional methods, CNNs possess the ability to extract complex hierarchical features directly from raw imaging data, demonstrating remarkable proficiency in recognising anatomical landmarks and interpreting radiographic patterns. In the dental field, deep learning has already shown promise across a spectrum of diagnostic tasks, from caries detection and periodontal assessment to the analysis of complex craniofacial structures [8,9]. Specific to skeletal maturation, several recent studies have explored AI-driven CVM staging. For instance, some researchers have employed single-stage classification models that analyse the entire cephalogram; however, these often suffer from "background noise" in the image that is irrelevant to the vertebrae themselves. Others have looked at multi-stage approaches, but the seamless integration of precise vertebral segmentation with accurate stage classification remains a challenge in the quest for clinical-grade accuracy. While these technological strides are encouraging, the clinical adoption of AI systems hinges on their ability to match the performance of expert clinicians under rigorous evaluation. The current study was designed to bridge the gap between raw radiographic data and clinical diagnosis by developing a fully automated, two-stage deep learning framework. The first stage focuses on the precise detection and segmentation of the C2-C4 vertebrae to isolate the regions of interest. The second stage then classifies these isolated regions into the standard CVM stages (CS1–CS6). We evaluated the performance of this framework using standardised detection and classification metrics, benchmarked against expert-labelled reference standards, to determine its viability as a reliable clinical decision-support tool.

Results

Performance of the Cervical Vertebrae Detection Model

In the initial stage, we compared various object detection architectures, including Cascade R-CNN and Mask R-CNN. The Mask R-CNN framework, integrated with a ResNet-50-FPN backbone, demonstrated

the most robust performance for the automated localisation of C2, C3, and C4 vertebrae and was thus selected for the final pipeline.

The model's efficacy was quantified using standard computer vision metrics. After a training regimen of 3,300 iterations, the model attained a mean Average Precision (mAP@50) of 0.85, indicating a high level of localisation reliability at an IoU threshold of 0.5 (Fig. 1).

Detailed analysis of the IoU thresholds revealed that at a moderate spatial overlap (IoU = 0.5), the model achieved an AP of 0.96. However, under more stringent localisation constraints (IoU = 0.75), the AP adjusted to 0.66, reflecting the inherent complexity of precise boundary delineation in radiographic imaging (Fig. 2). The overall recall was 0.62, signifying a stable detection rate across the heterogeneous dataset. The confusion matrix (Fig. 3) further confirms high discriminatory power, particularly for the C4 vertebra, with minimal cross-classification between adjacent vertebral levels. To further evaluate the reliability of the detection model's output, Cohen's Kappa coefficient, the coefficient yielded a value of approximately $\kappa \approx 0.513$. This result indicates a moderate to good agreement between the model's predictions and the ground truth for vertebrae localization, suggesting that the detection performance surpasses random chance.

Performance of Stage II: CVM Stage Classification

In the second stage, the 756 successfully segmented triplets (C2–C4) were processed through the EfficientNet-B3 classifier. The diagnostic performance across the six maturation stages (CS1–CS6) is detailed in the confusion matrix (Fig. 4) and summarised in Table 1.

Table 1

Stage	Precision	Recall	F1-Score
CS1	61%	54%	57%
CS2	42%	53%	47%
CS3	32%	30%	31%
CS4	37%	34%	35%
CS5	45%	34%	39%
CS6	62%	73%	67%

The model demonstrated its peak performance in identifying CS6 (Late Maturation), achieving an accuracy of 73% (38/52 samples). The early stages, CS1 and CS2, showed moderate-to-high classification reliability, both exceeding the 50% accuracy threshold.

Conversely, the intermediate stage (CS3) proved to be the most challenging for the framework, yielding a classification accuracy of 33%. Most misclassifications for CS3 were attributed to Stage CS2, reflecting

the clinical difficulty in detecting the initial onset of the concavity This performance dip is primarily attributed to the "morphological transition" phase; the subtle changes in the concavity of the lower borders and the transition from trapezoidal to rectangular shapes in C3 and C4 often lead to overlapping features between CS2, CS3, and CS4.

The Cohen's Kappa coefficient, calculated to evaluate the agreement between the model's predictions and the ground, yielded a value of approximately 0.37200 and this indicates that the model's performance is better than random chance, but there is still significant room for improvement, and the model is not yet fully capable of accurately distinguishing between all classes.This result can be attributed to several factors:

- Morphological Complexity: The inherent difficulty in distinguishing between mid-maturation stages (particularly CS3) due to subtle transitional features and overlapping anatomical characteristics poses a significant challenge for accurate classification.
- Stage I Dependency: The quality of the input data (756 C2-C4 triplets) extracted from the initial vertebrae detection stage directly impacts the classification accuracy. Potential errors in localization or segmentation from Stage I can lead to misclassifications in Stage II.
- Data Distribution: The nature and distribution of samples within the 756-instance dataset, especially for stages with fewer samples, may have influenced the model's ability to generalize and achieve higher agreement.

In summary, the Kappa of 0.37200 reflects the challenges in accurately classifying CVM stages, influenced by morphological complexities and the quality of upstream detection data.

Discussion

The integration of artificial intelligence (AI) into orthodontic diagnostics has emerged as a transformative approach to mitigating the inherent subjectivity of skeletal maturity assessment. The present study successfully developed and validated a fully automated, two-stage deep learning framework for CVM assessment. Our clinical pipeline achieved two primary milestones: first, a highly precise localisation of the C2–C4 vertebrae with an AP of 0.96 (at IoU = 0.5), and second, a robust classification of skeletal maturity stages using an EfficientNet-B3 architecture, reaching a peak accuracy of 73% for the final maturation stage (CS6). While the model demonstrated high reliability in identifying polar stages (CS1, CS2, and CS6), it also highlighted the inherent diagnostic complexity of the transitional CS3 stage, which yielded a lower accuracy of 33%. These results underscore the potential of hierarchical deep learning models to standardise orthodontic growth analysis while maintaining high anatomical transparency.

Comparative Performance and Architectural Advantages

Methodologically, our framework addresses the limitations of prior "black-box" approaches that perform direct classification on entire radiographs. For instance, while Seo et al. and Li et al. reported varying accuracies using direct classification [10,11], these models often lack anatomical transparency. By implementing an explicit detection and segmentation stage, our model ensures that the feature extraction process is strictly confined to the C2–C4 vertebral bodies, effectively filtering out non-contributory craniofacial structures—a frequent source of noise in single-stage models [12,13]. Our late-fusion architecture independently encodes C2–C4 vertebrae to preserve local morphological nuances while capturing inter-vertebral spatial correlations. This hierarchical approach ensures a more robust and physiologically aligned representation of skeletal maturation compared to conventional single-stream models.

Furthermore, while semi-automated methods like those proposed by Kavousinejad et al. [14] achieve high accuracy, they remain dependent on manual landmarking, which is time-consuming and prone to inter-observer bias. In contrast, our end-to-end framework offers superior clinical scalability, eliminating manual intervention while maintaining high fidelity in anatomical localisation.

The Challenge of Transitional Stages (CS3)

A notable finding in our study was the performance disparity between polar and intermediate maturation stages. The model achieved its highest sensitivity in CS6 (73%), whereas CS3 exhibited the lowest accuracy (33%). This phenomenon is consistent with the findings of Jiang et al. and likely reflects the "morphological continuum" of puberty. During the CS3 stage, the transition from a flat to a concave lower border in C3 and C4 is often subtle and lacks a discrete threshold, posing significant challenges even for experienced clinicians [15,16]. The high localisation precision of our Stage I model suggests that this misclassification is rooted in the morphological overlap between classes rather than a failure in feature detection.

Clinical Implications and Robustness

To bridge the gap between experimental models and "real-world" clinical practice, we purposefully integrated 187 noisy and low-quality images into our training pipeline. This strategy, combined with the use of the Detectron2 framework for standardised pre-processing, enhances the model’s generalizability across different radiographic hardware and exposure settings. From a clinical standpoint, this tool could serve as a reliable "second opinion" for orthodontists, standardising the timing of functional appliance therapy and orthognathic surgery.

Limitations and Future Directions

Despite the promising results, certain limitations warrant acknowledgement. The moderate recall (0.62) in the detection stage indicates that some vertebral structures were missed, potentially due to severe

postural variations or suboptimal contrast in the archives. Additionally, the inherent class imbalance in pediatric populations remains a challenge for deep learning models.

Future research should focus on:

Multi-branch Fusion Architectures: Leveraging the independent morphological signals of C2, C3, and C4 more effectively through advanced feature fusion.

Dataset Expansion: Utilizing multi-center datasets to further mitigate class imbalance and improve performance during the transitional CS3–CS4 phases.

Conclusion

The two-stage deep learning framework developed provides a robust and fully automated solution for CVM assessment, eliminating the subjectivity of manual methods. By coupling high-precision vertebral segmentation with an EfficientNet-based classifier, the model achieves high diagnostic reliability, particularly in identifying polar maturation stages. Despite the challenges posed by the morphological nuances of transitional phases (CS3), the system's resilience to "real-world" image noise demonstrates its potential as a scalable auxiliary tool in clinical orthodontics. This pipeline offers a standardized approach to skeletal growth assessment, facilitating more precise timing for growth-related interventions.

Materials and Methods

Sampling Method and Sample Size Determination

In this retrospective cross-sectional study, a comprehensive dataset of lateral cephalometric radiographs (LCRs) of individuals aged 7 - 18 years was retrieved from the digital archives of two private orthodontic clinics in Tehran and Yazd, Iran, as well as the Department of Orthodontics, Faculty of Dentistry, Shahid Sadoughi University of Medical Sciences, Yazd, Iran spanning from spanning from January 2018 to December 2023 .The study population comprised. Utilizing routine clinical archives ensured a diverse representation of radiographic images obtained under standard diagnostic conditions. To ensure the integrity of the vertebral morphological analysis, the following exclusion criteria were applied:

- Presence of systemic diseases, endocrine disorders, or developmental delays known to impair skeletal growth or bone maturation.
- Congenital or acquired craniofacial anomalies or cervical spine deformities.
- Suboptimal image quality, including artefacts, positioning errors, or inadequate visualization of the second (C2), third (C3), and fourth (C4) cervical vertebrae.

Following the initial screening, a total of 1,390 high-quality LCRs were included. This sample size was determined to meet the high data-dimensionality requirements of deep learning architectures, ensuring robust feature extraction and model generalizability.

Ethics approval and consent to participate

The study protocol was strictly conducted in accordance with the Declaration of Helsinki and received formal approval from the Ethics Committee of Shahid Sadoughi University of Medical Sciences (Protocol No: IR.SSU.DENTISTRY.REC.1403.074). Given the retrospective nature of the study and the use of de-identified, anonymized data, the requirement for written informed consent was waived by the institutional review board.

Data Pre-processing and Quality Control

Image Standardisation and Curation

The initial dataset consisted of 1,390 lateral cephalometric radiographs (LCRs) originally captured in Digital Imaging and Communications in Medicine (DICOM) format. To streamline the deep learning workflow, these images were converted to Portable Network Graphics (PNG) format. **The raw dataset exhibited significant heterogeneity in resolution, with widths ranging from 568 to 2,144 pixels and heights from 570 to 2,600 pixels.**

A rigorous quality control (QC) protocol was implemented to ensure the reliability of the training data. Following an expert review, 288 radiographs were excluded due to excessive noise, motion artifacts, or failure to meet the anatomical inclusion criteria (e.g., inadequate visualization of C2–C4). This resulted in a refined primary dataset of 1,102 images for subsequent analysis.

Class Distribution and Imbalance

Analysis of the CVM stages (CS1–CS6) revealed a significant class imbalance, with CS3 being the least represented category (Figure 5). To mitigate the risk of biased learning toward majority classes, we implemented specific strategies, including:

- Cost-sensitive learning (class weighting).
- Targeted data augmentation for underrepresented stages.

Exploratory Data Analysis (EDA)

Comprehensive EDA was performed to evaluate the geometric and photometric characteristics of the dataset (Figure 6). The distribution of image dimensions (width, height, and aspect ratio) displayed multimodal patterns, suggesting that dimensional standardisation was not a prerequisite for the feature extraction layers. Photometric analysis revealed that the mean brightness and pixel intensity followed a Gaussian distribution centered on mid-gray levels, indicating consistent exposure across the archives.

Feature Complexity and Noise Strategy

To quantify image detail, edge density was calculated for the entire dataset (Figure 7). The majority of the images fell within the 0.07–0.18 range. Images at the lower tail of the distribution (edge density < 0.04) were identified as low-detail samples.

Furthermore, to enhance the model's robustness against real-world clinical variations, a strategic decision was made regarding noisy data. Instead of total exclusion, 187 images with controlled levels of noise were specifically retained and integrated into the data augmentation pipeline. This approach was designed to minimize overfitting and ensure that the framework maintains high diagnostic performance even when processing suboptimal, "non-ideal" clinical radiographs (Figure 5).

Model Design

Proposed AI Framework

The proposed diagnostic pipeline follows a two-stage deep learning architecture: (1) automated detection and instance segmentation of the C2, C3, and C4 cervical vertebrae from LCRs, and (2) multi-class classification of skeletal maturity stages based on the localised vertebral regions.

Dataset Preparation and Stratification

The experimental dataset was categorised into two subsets: noisy (n = 187) and clean (n = 915) images, as detailed in Figure 8. To maintain statistical balance while ensuring the model's robustness against real-world artefacts, we adopted a differential splitting strategy. For the noisy subset, an 80/19/1% split was used for training, validation, and testing, respectively. The clean subset followed a standard 70/20/10% distribution. This partitioning ensures that the model is exposed to a high volume of suboptimal data during training, thereby improving its generalisation across diverse clinical environments (Figure 9). **Stage I: Vertebral Detection and Instance Segmentation**

Ground Truth Generation: Manual annotations were performed by orthodontic specialists using a web-based tool. Polygonal contours were delineated for C2, C3, and C4 to capture precise morphological boundaries.

Architecture and Transfer Learning: We implemented a Mask R-CNN architecture with a ResNet-50 backbone and a Feature Pyramid Network (FPN) to facilitate multi-scale feature extraction. To leverage prior knowledge, transfer learning was employed by initialising the network with weights pre-trained on the COCO dataset. During fine-tuning, the initial convolutional layers were frozen to retain low-level generic features, while deeper layers were optimised to learn the specific morphology of cervical vertebrae.

Inference and Refinement: During the inference phase, a confidence threshold of 0.5 and a Non-Maximum Suppression (NMS) threshold of 0.6 were applied to filter out low-probability or redundant

detections. Only radiographs where all three target vertebrae (C2–C4) were detected with an Intersection-over-Union (IoU) > 0.5 were retained. This rigorous filtering resulted in a curated set of 756 radiographs for the second stage (Figures 9, 10).

3.3. Stage II: Skeletal Maturity Classification

Localised Pre-processing: The predicted masks from Stage I were used to crop the C2, C3, and C4 regions. Each vertebral crop was normalised and resized to a fixed resolution of 224 × 224 pixels. The final classification dataset (n = 756) was divided into training (70%) and testing (30%) sets using stratified sampling to preserve the distribution of CVM stages (CS1–CS6) across both cohorts (Figure 11).

Classification Network and Late-Fusion Strategy: We developed a triple-input classification network based on EfficientNet-B3. The architecture was modified to accept single-channel grayscale inputs, with weights initialised from ImageNet.

Data Augmentation: To prevent overfitting and simulate clinical variations, independent stochastic augmentations were applied to each vertebral crop:

Geometric: Random rotations (±15∘), horizontal flipping, and translations (±10%).

Photometric: Isotropic scaling (0.9–1.1) and pixel intensity normalisation (μ=128, σ=64).

Feature Fusion and Output: Each vertebra (C2, C3, and C4) was processed through a dedicated EfficientNet-B3 branch. Following the Global Average Pooling and flattening layers, three 1,536-dimensional feature vectors were generated. We implemented a late-fusion strategy by concatenating these vectors into a comprehensive 4,608-dimensional representation. This fused vector was then passed through a dropout layer and a fully connected layer with a Softmax activation function, yielding the final class probabilities for stages CS1 through CS6. This approach allows the model to learn both the individual morphological changes of each vertebra and their collective spatial relationships.

Declarations

Data Availability

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request, subject to ethical considerations and institutional data privacy policies

Author Contributions

Conceptualization and study design: MJM

Dataset preparation and radiographic annotation: MJM, HAA, AMF, YS, MEG

Development of the two-stage deep learning framework: MZE, HD, AK

Model training, testing, and performance evaluation: MZE

Statistical analysis: HD

Clinical interpretation of results: MJM, HAA, MZE

Writing – original draft: AK

Writing – review & editing: MJM, MZE

Supervision and final approval: MJM

Additional Information

Funding Declaration:

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

Competing interests:

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

References

1. Hägg, U. & Taranger, J. Skeletal stages of the hand and wrist as indicators of the pubertal growth spurt. *Acta Odontologica Scandinavica* **38**, 187-200 (1980).
2. Fishman, L. S. Radiographic evaluation of skeletal maturation. *The Angle Orthodontist* **52**, 88-112 (1982).
3. Cha, K.-S. Skeletal changes of maxillary protraction in patients exhibiting skeletal class III malocclusion: a comparison of three skeletal maturation groups. *The Angle Orthodontist* **73**, 26-35 (2003).
4. Ngan, P. Early treatment of Class III malocclusion: is it worth the burden? *American journal of orthodontics and dentofacial orthopedics* **129**, S82-S85 (2006).
5. Kucukkeles, N., Acar, A., Biren, S. & Arun, T. Comparisons between cervical vertebrae and hand-wrist maturation for the assessment of skeletal maturity. *The Journal of clinical pediatric dentistry* **24**, 47-52 (1999).
6. Gabriel, D. B. *et al.* Cervical vertebrae maturation method: poor reproducibility. *American Journal of Orthodontics and Dentofacial Orthopedics* **136**, 478. e471-478. e477 (2009).
7. Hassel, B. & Farman, A. G. Skeletal maturation evaluation using cervical vertebrae. *American Journal of Orthodontics and Dentofacial Orthopedics* **107**, 58-66 (1995).

8. Arık, S. Ö., Ibragimov, B. & Xing, L. Fully automated quantitative cephalometry using convolutional neural networks. *Journal of Medical Imaging* **4**, 014501-014501 (2017).
9. Lee, J.-H., Kim, D.-H., Jeong, S.-N. & Choi, S.-H. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *Journal of dentistry* **77**, 106-111 (2018).
10. Seo, H., Hwang, J., Jeong, T. & Shin, J. Comparison of deep learning models for cervical vertebral maturation stage classification on lateral cephalometric radiographs. *Journal of Clinical Medicine* **10**, 3591 (2021).
11. Li, H. *et al.* Convolutional neural network-based automatic cervical vertebral maturation classification method. *Dentomaxillofacial Radiology* **51**, 20220070 (2022).
12. Zhou, J. *et al.* Development of an artificial intelligence system for the automatic evaluation of cervical vertebral maturation status. *Diagnostics* **11**, 2200 (2021).
13. Kim, E.-G. *et al.* Estimating cervical vertebral maturation with a lateral cephalogram using the convolutional neural network. *Journal of Clinical Medicine* **10**, 5400 (2021).
14. Kavousinejad, S., Ebadifar, A., Tehranchi, A., Zakermashhadi, F. & Dalaie, K. Determination of cervical vertebral maturation using machine learning in lateral cephalograms. *Journal of Dental Research, Dental Clinics, Dental Prospects* **18**, 232 (2024).
15. Jiang, F. *et al.* Deep learning based quantitative cervical vertebral maturation analysis. *Head & Face Medicine* **21**, 20 (2025).
16. Amasya, H., Yildirim, D., Aydogan, T., Kemaloglu, N. & Orhan, K. Cervical vertebral maturation assessment on lateral cephalometric radiographs using artificial intelligence: comparison of machine learning classifier models. *Dentomaxillofacial Radiology* **49**, 20190441 (2020).

Figures

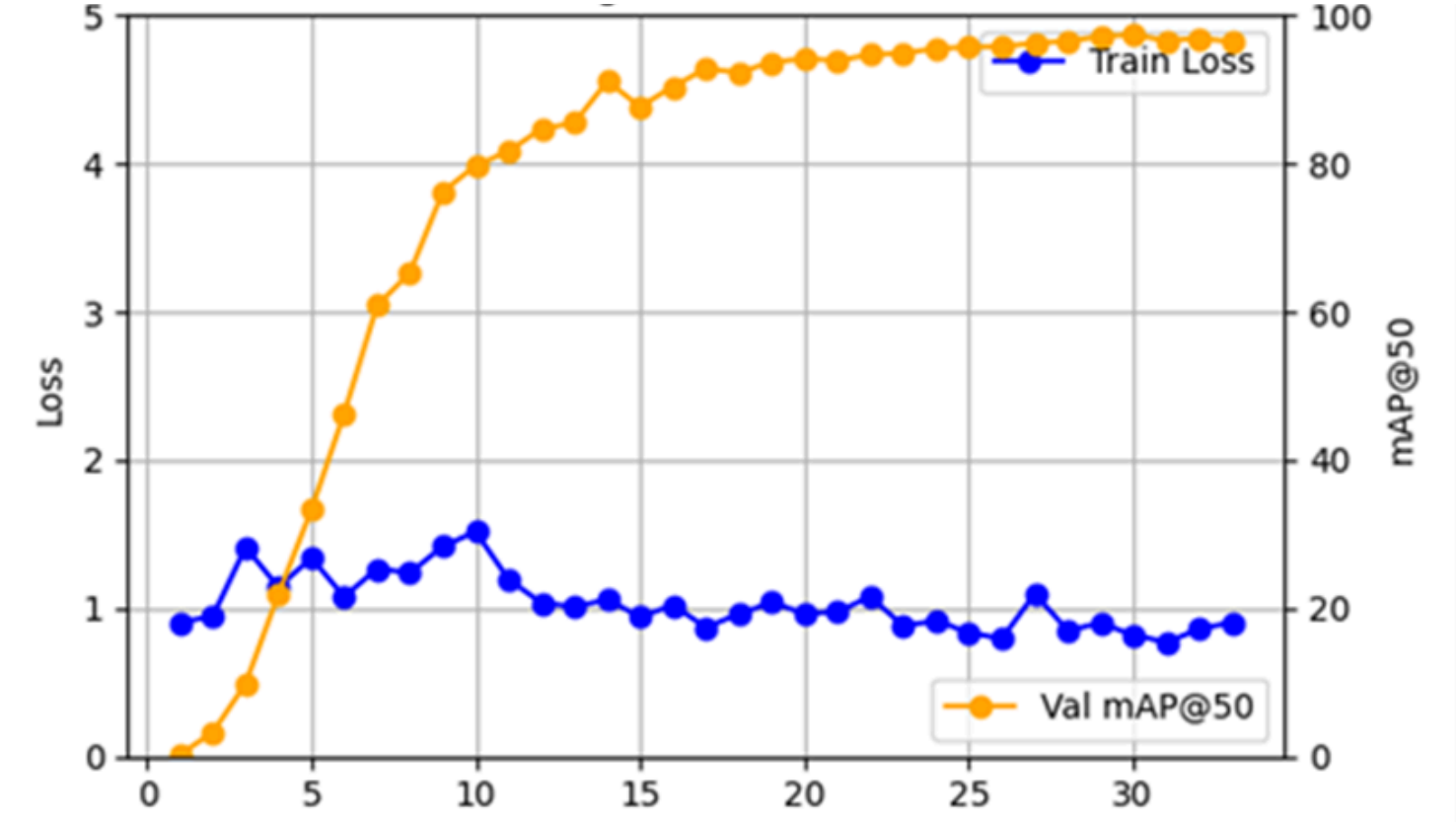

Figure 1

Training loss and test mean Average Precision (mAP@0.50) for the cervical vertebrae detection model.

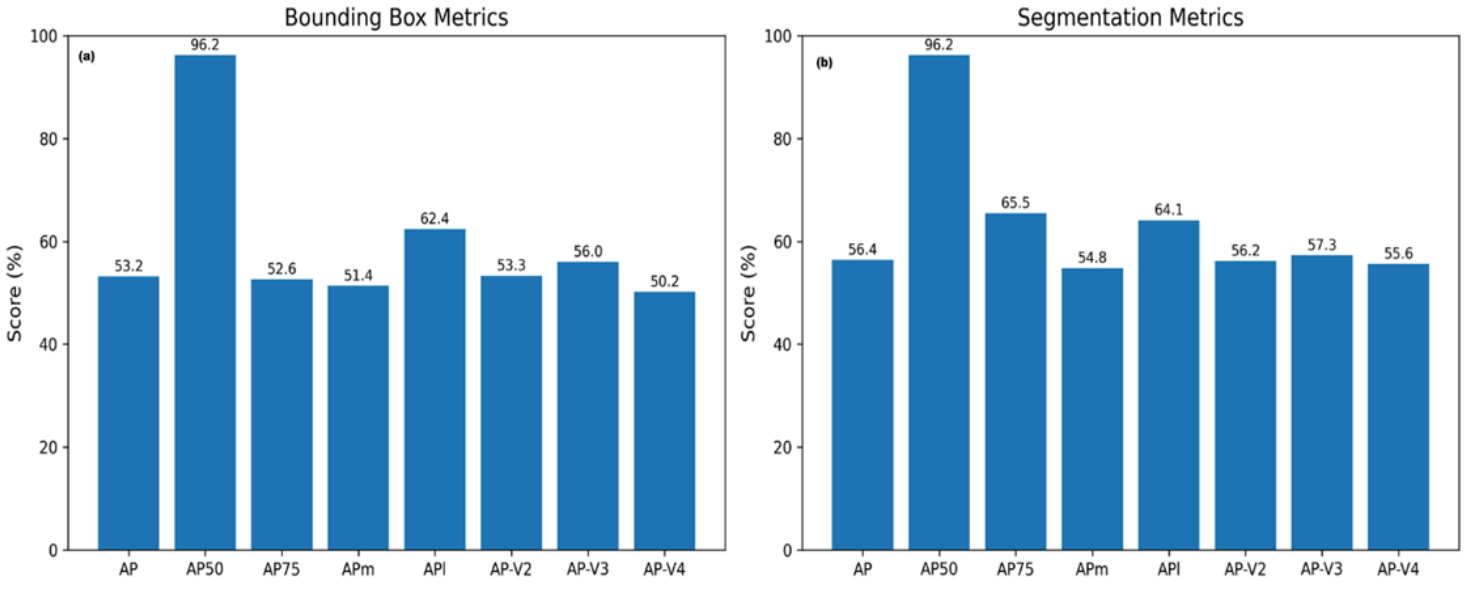

Figure 2

Performance metrics of the cervical vertebrae detection model: (a) bounding box detection results and (b) segmentation results, reported using Average Precision (AP) at different IoU thresholds and object scales

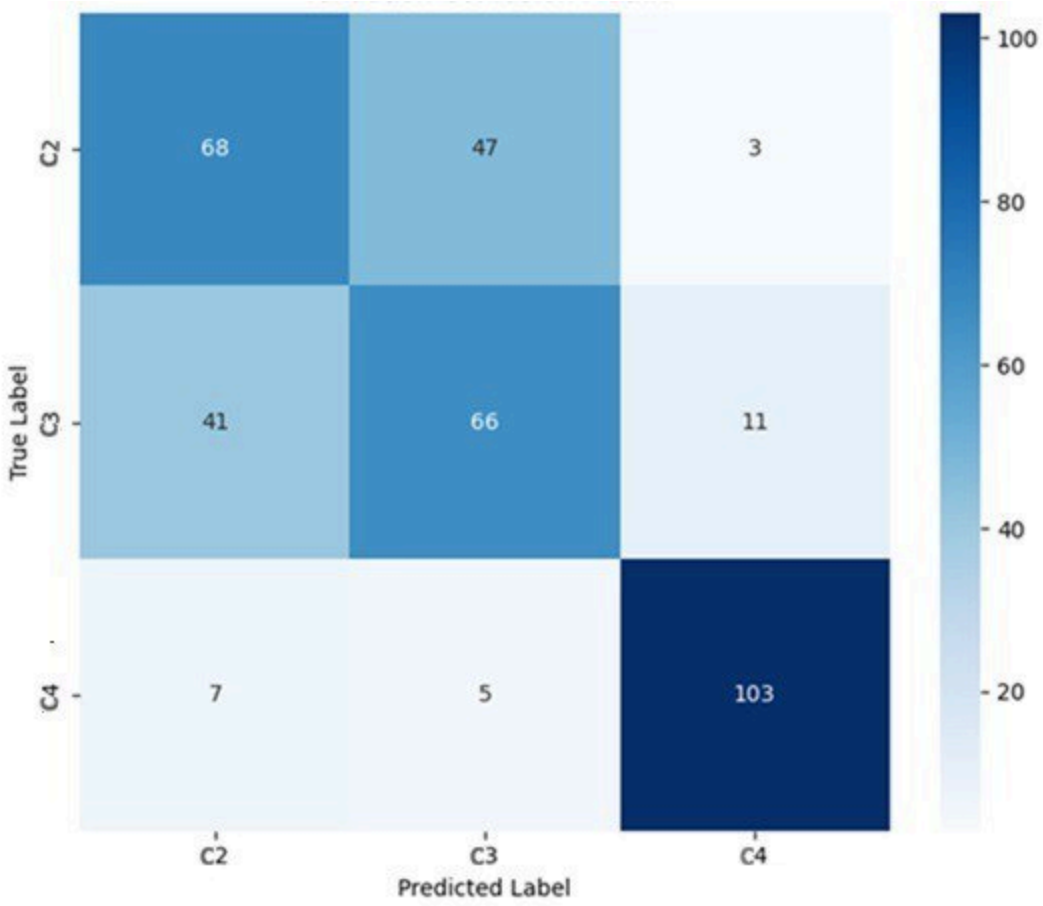

Figure 3

Confusion matrix of the cervical vertebrae detection model showing the classification performance for vertebrae C2, C3, and C4.

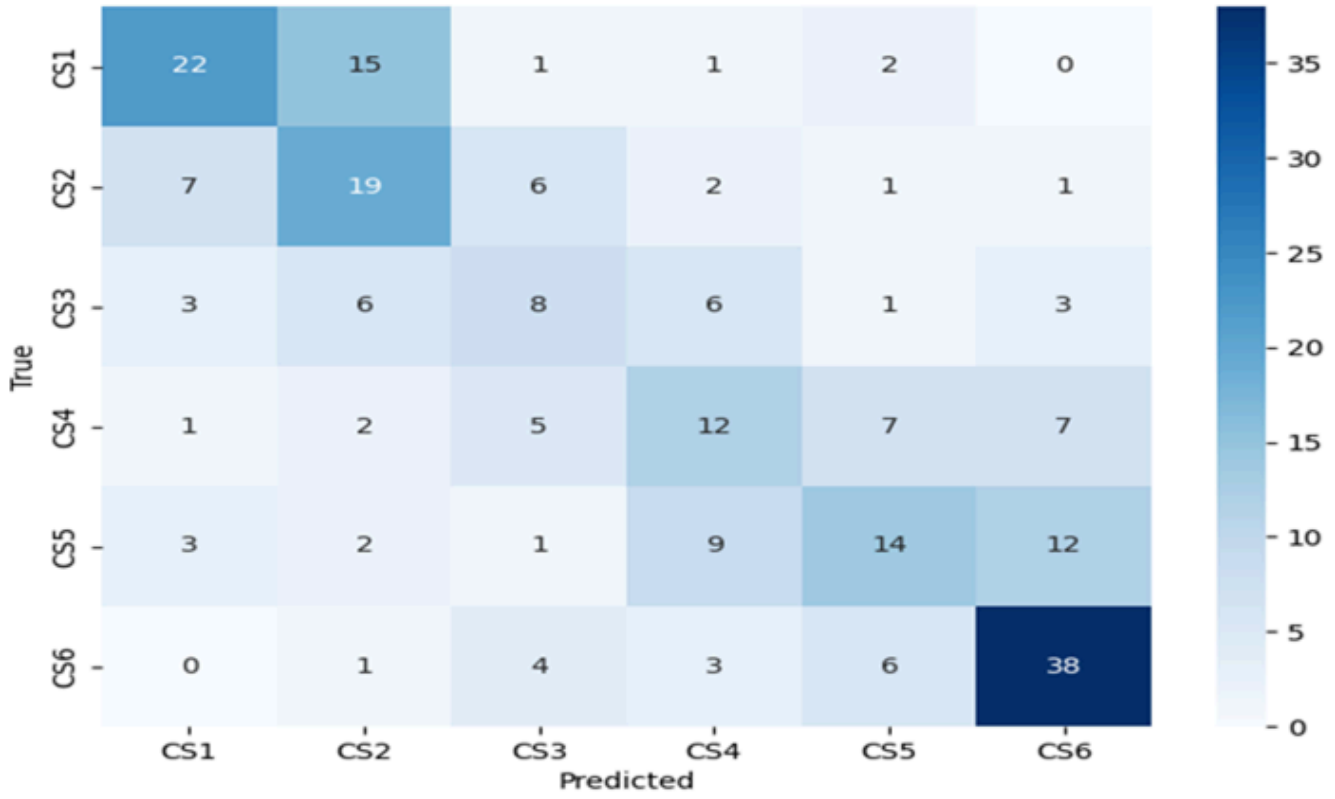

Figure 4

Confusion matrix resulting from classifier model

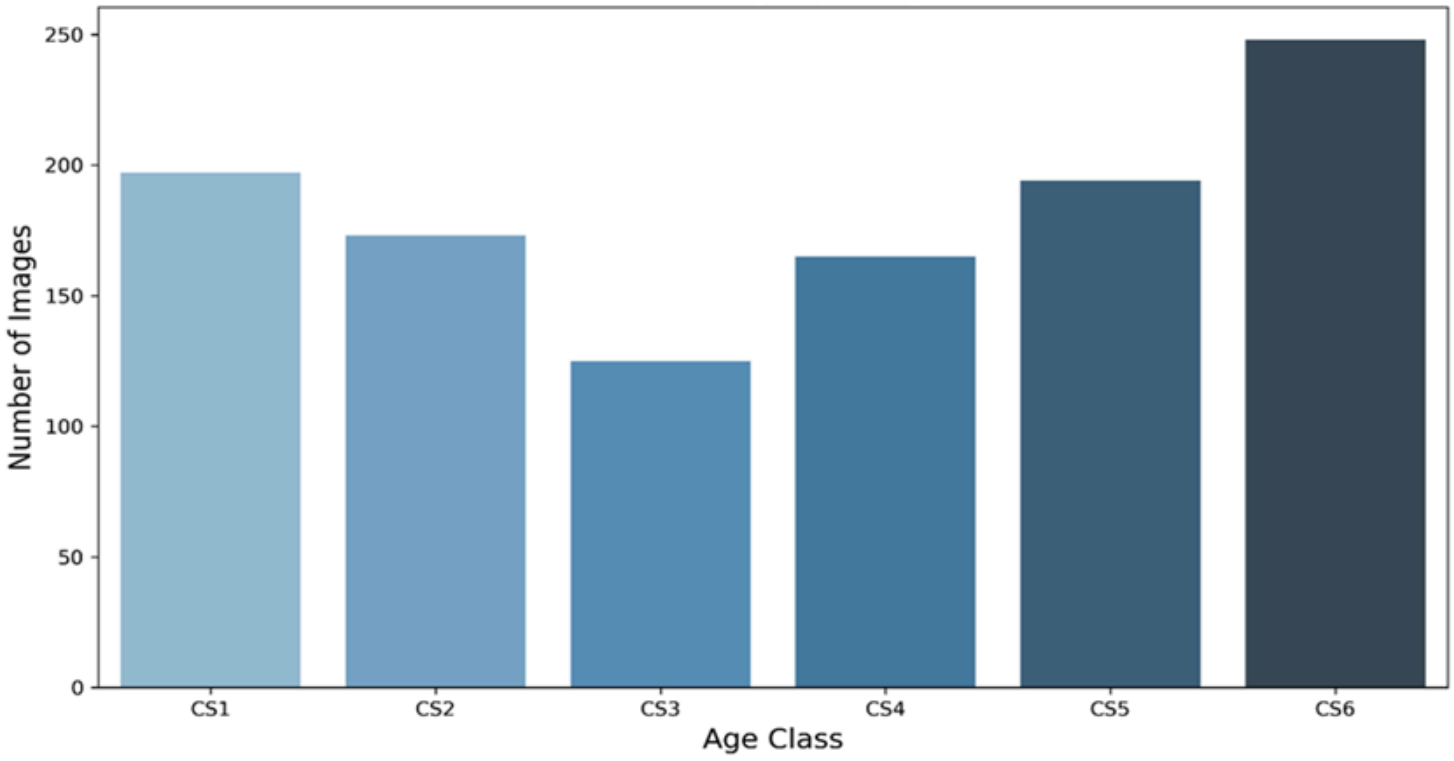

Figure 5

Distribution of cervical vertebral maturation stages (CS1–CS6)

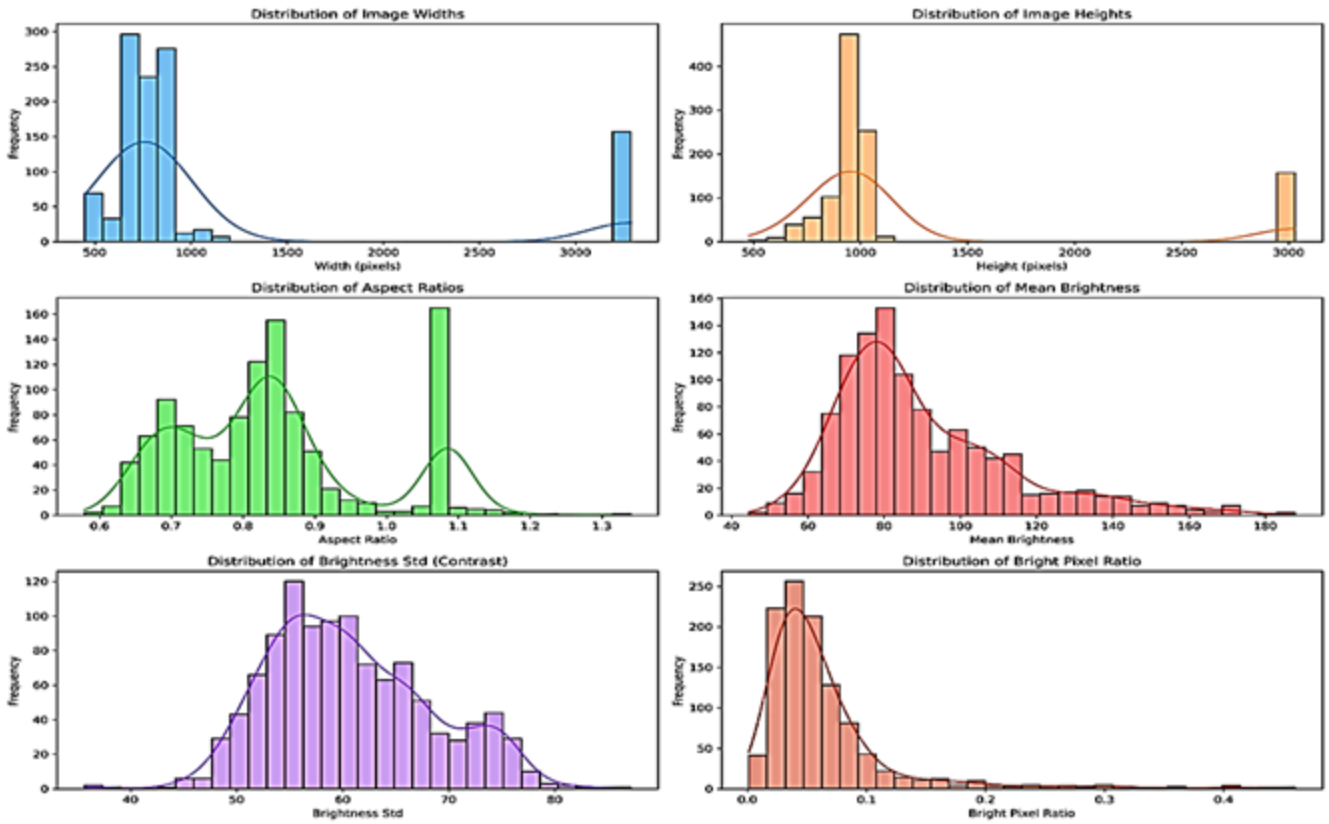

Figure 6

Exploratory data analysis of geometric and photometric image characteristics.

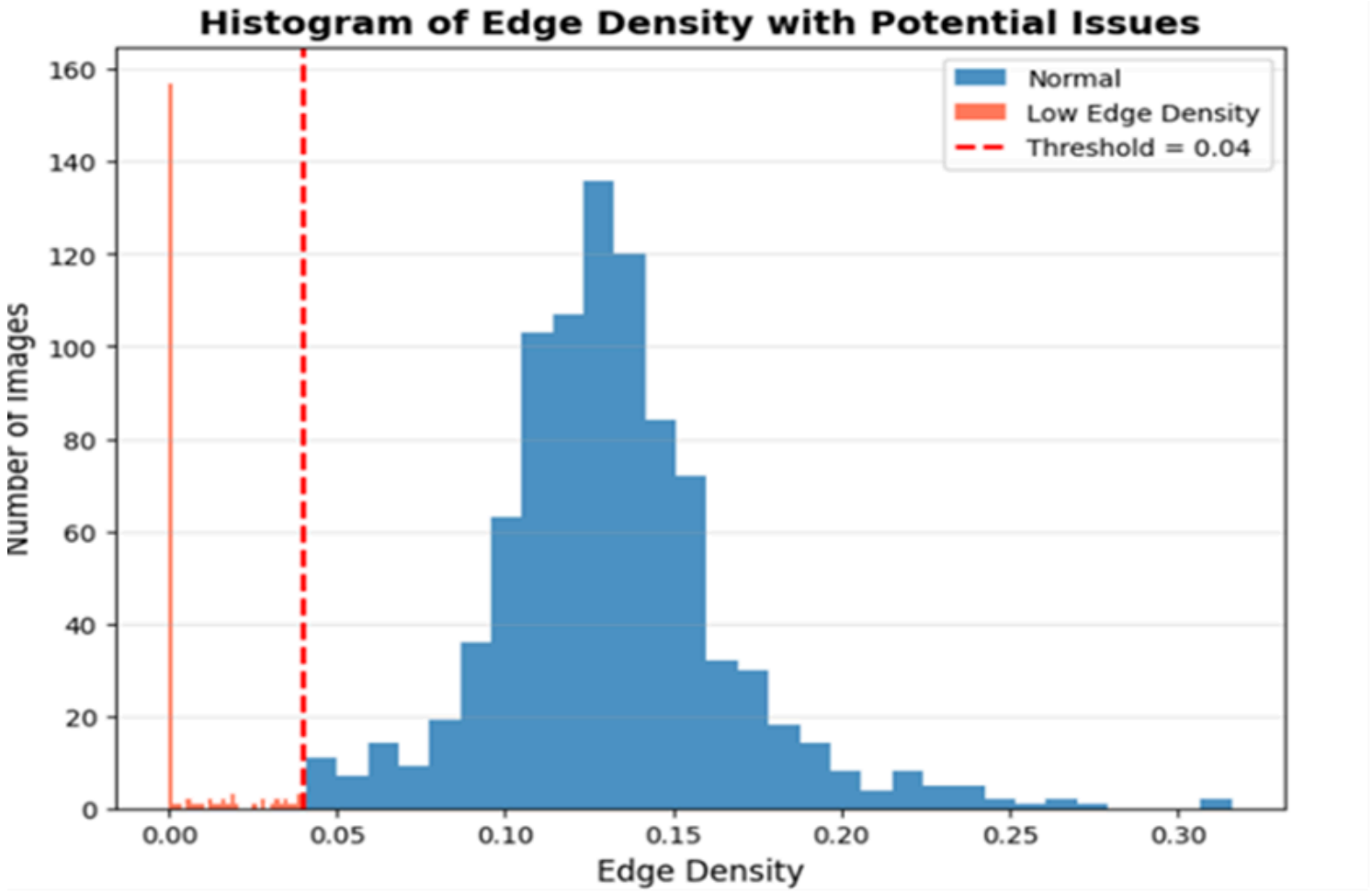

Figure 7

Edge density distribution analysis revealing primary cluster (0.07–0.18 range) and low-detail tail (<0.04) used for quality flagging. Majority of images fall within optimal edge density range for vertebra detection

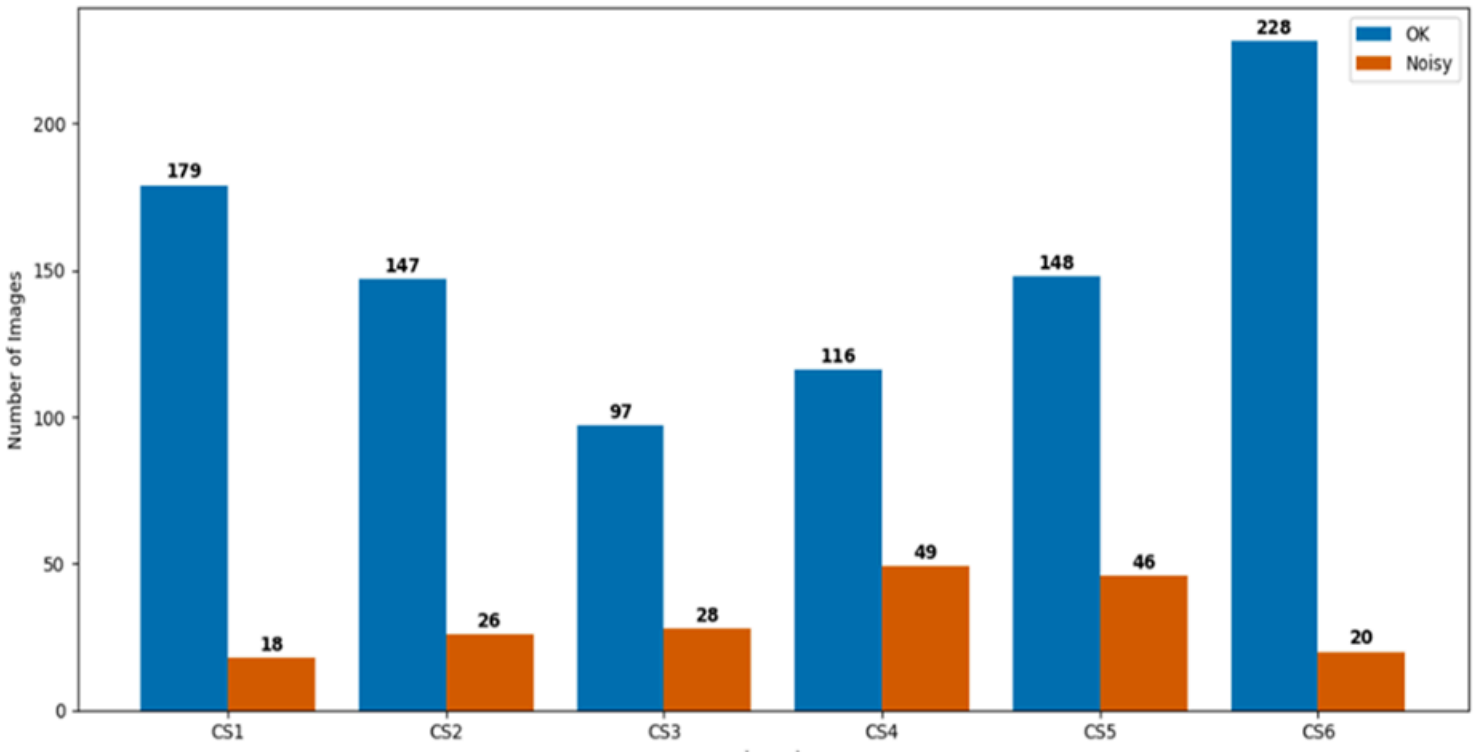

Figure 8

Distribution of noisy data in the dataset

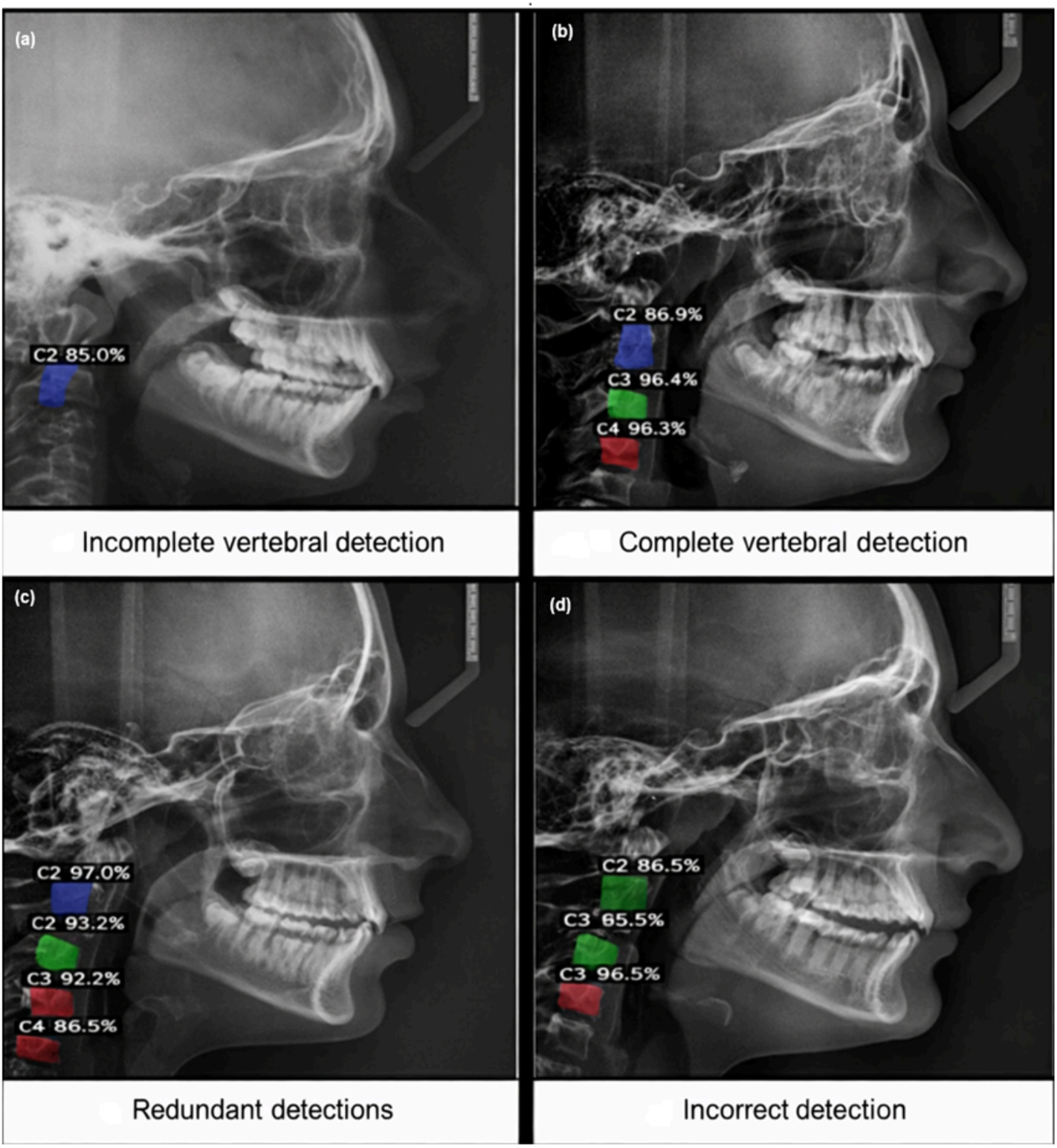

Figure 9

Examples of cervical vertebrae detection outcomes using Detectron2.(a) Incomplete detection of cervical vertebrae, (b) complete detection of C2–C4, (c) redundant detections with overlapping masks, and (d) incorrect detection. Detected vertebrae are highlighted using color-coded masks, and confidence scores displayed for each prediction.

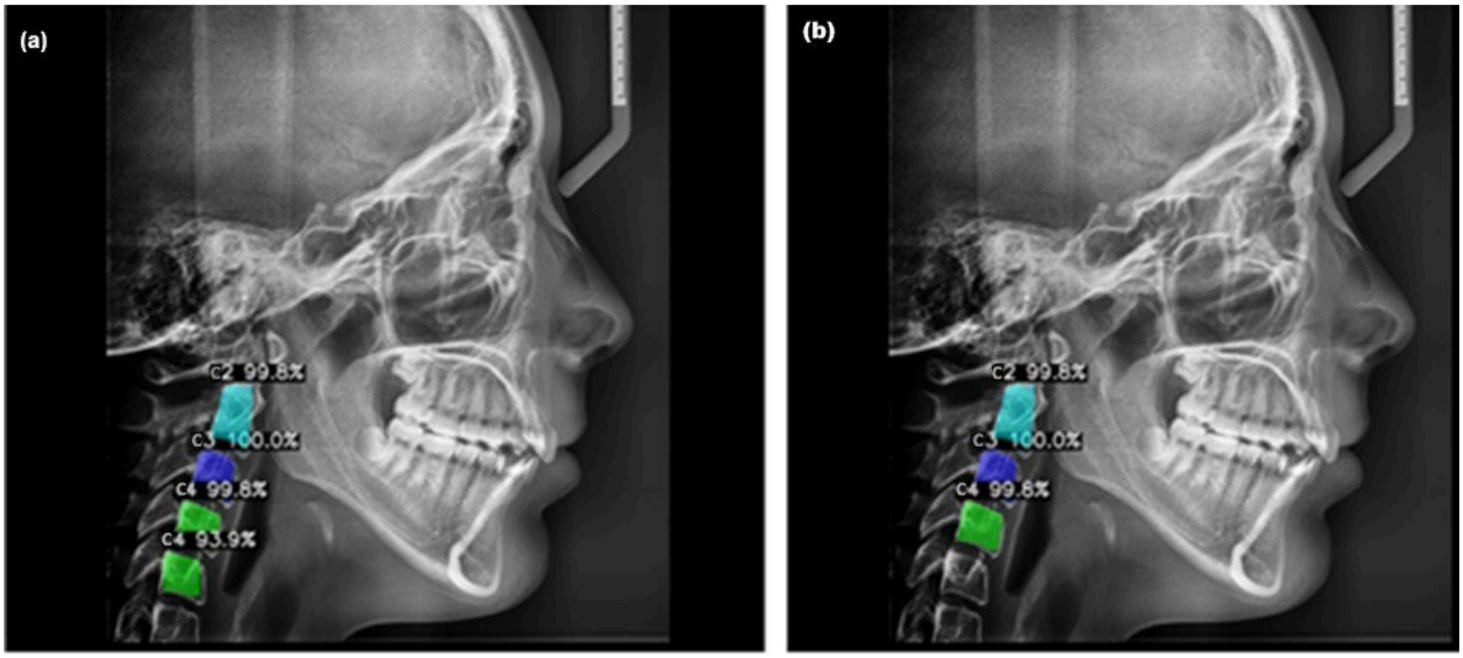

Figure 10

Effect of detection optimization on cervical vertebrae segmentation results. (a) Initial detection results before post-processing, showing redundant or missing vertebral predictions. (b) Optimized detection after filtering duplicate masks and retaining predictions with higher confidence scores. Detected cervical vertebrae (C2–C4) are highlighted using color-coded masks with corresponding confidence values.

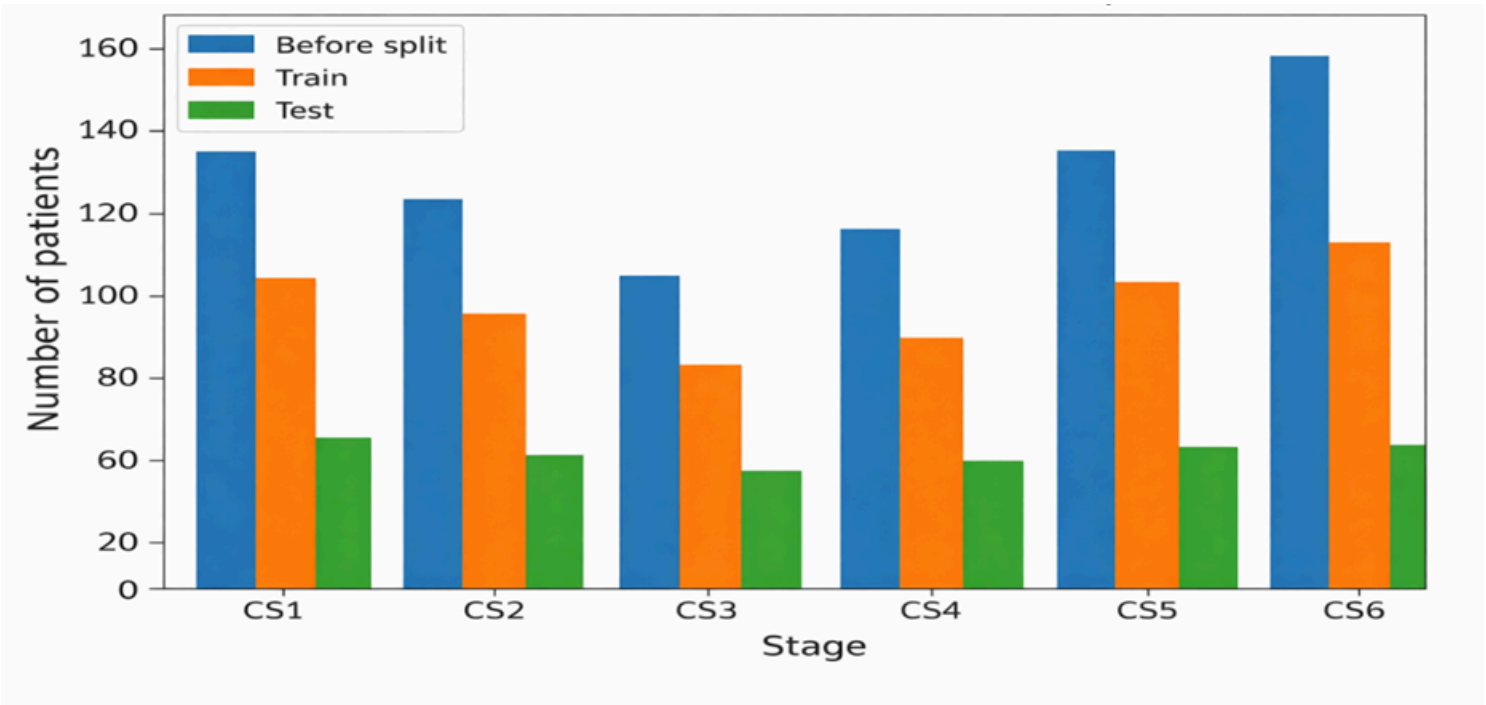

Figure 11

Distribution of samples obtained from the cervical vertebrae detection and segmentation stage (C2–C4) across different skeletal maturity classes (CS1–CS6), and their allocation into training and test sets for skeletal age classification.