Cervical Vertebral Maturation Staging from a Continuous Perspective

Kush Mahajan[1*], Dhruv Gupta[1*], Gagneet Singh[2], Vinay Soni[3], Manojkumar Jaiswal[3], and Sukrit Gupta[1#]

[1] Indian Institute of Technology, Ropar, India
[2] University Institute of Engineering and Technology, Panjab University, Chandigarh, India
[3] Post Graduate Institute of Medical Education and Research, Chandigarh, India

Abstract. Cervical vertebral maturation (CVM) staging guides orthodontic treatment and is conventionally formulated as a set of six discrete classes. In practice, skeletal development progresses gradually, and many lateral cephalographic radiographs exhibit transitional morphology between adjacent stages, resulting in well-documented, substantial inter-expert variability in CVM stage prediction. Prior Deep Learning (DL) work treats CVM as standard multi-class classification despite its gradual nature, using exact-match accuracy that ignores clinically acceptable Tolerance within ± 1 stage (Tol-1). We introduce the Ordinal RAnking of Cervical vertebrae maturation using deep LEarning (ORACLE) framework and reformulate CVM assessment as a continuous learning task modeling adjacent-stage relationships. A dataset comprising 1,248 lateral cephalograms (LCs) was compiled with expert-assigned stage labels; 569 with expert-verified C2-C4 vertebrae polygon masks. Across five stratified folds, the framework achieved 62.82% exact accuracy and 96.07% accuracy within Tol-1, confirming that adjacent-stage errors dominate, matching clinical expectation. On a separate set of 113 lateral cephalograms, we observed a similar trend of modest exact agreement but high adjacent-stage agreement with clinicians of varying experience. The ORACLE framework achieved quadratic weighted κ scores of 0.88 and 0.79 against two senior experts, demonstrating performance comparable to inter-expert agreement (κ=0.74). These findings demonstrate that CVM staging reflects a gradual developmental continuum not fully captured by rigid multi-class evaluation frameworks, supporting continuous-stage prediction with tolerance-based metrics as a clinically better-aligned framework for automated CVM assessment. The dataset is available upon request through the data access form at Link, and the source code is publicly available on GitHub.

Keywords: Cervical vertebral maturation, continuous learning, tolerance-based evaluation, skeletal maturity evaluation, inter-observer variability.

* These authors contributed equally to this work.
# Corresponding author: sukrit@iitrpr.ac.in

1 Introduction

Evaluation of skeletal maturation is essential in orthodontic treatment planning, particularly for determining the optimal timing of growth modification. Cervical Vertebral Maturation (CVM) method leverages C2-C4 vertebral concavity/shape evolution on lateral cephalograms (LCs) to stage skeletal maturity into 6 discrete phases, guiding optimal timing for growth modulation therapies (Class II/III malocclusions) where stage timing errors alter treatment outcomes [1,2]. Although defined as six discrete stages, skeletal development progresses gradually. Transitional radiographs often exhibit characteristics of neighboring stages, contributing to inter- and intra-expert variability in CVM staging even among experienced clinicians with agreement particularly decreasing between adjacent stages and even the same clinician assigning different stages upon re-evaluation after a time interval [3,4,5,6]. CVM's continuous nature demands modeling beyond categorical cutoffs to capture clinically acceptable Tolerance within ± 1 stage (Tol-1).

Recent advances in Deep Learning (DL) have driven the development of an increasing number of fully automated CVM assessment systems. We reviewed studies that take a raw lateral cephalogram (LC) as input and directly output a CVM stage without requiring manual landmark annotation during inference; i.e. methods that require clinician-provided annotations at inference time were excluded from consideration. From Table 1, several consistent gaps characterize the prior literature. Gap 1: ROI vs pixel-precise C2-C4 contours needed for concavity / shape (clinical gold standard metrics). Gap 2: although CVM represents a continuous biological progression, existing approaches treat it as a standard multi-class classification problem. Gap 3: While a few studies report model-expert comparisons, agreement values are typically moderate. In addition, some works simplify the staging protocol by reducing six stages into three broader categories or modifying the staging scheme, which may partly contribute to their better performance.

To better align modeling with biological progression, we studied three complementary formulations: ordinal learning via COnsistent RAnk Logits (CORAL) loss [15], distribution-based optimization using the squared Earth Mover's Distance (EMD) loss [16], and direct regression-based stage estimation. While such approaches have demonstrated promise in various medical grading and estimation tasks [17,18,19], they remain underexplored in the context of CVM staging. We compiled a dataset of 1,248 LCs, including 569 images with expert-verified polygon annotations of C2–C4 vertebrae. A fully automated segmentation-to-stage pipeline, Ordinal RAnking of Cervical vertebrae maturation using deep LEarning (ORACLE) is developed and stage prediction is evaluated under the three formulations. In addition to conventional metrics, tolerance-based evaluation is performed, and inter-expert variability is measured on a separate held-out dataset involving clinicians with differing experience levels. Model-expert agreement is contextualized against this variability to assess clinical plausibility.

Table 1. Summary of fully automated DL approaches for CVM staging. 'Public Dataset' indicates whether the dataset is openly accessible. Extraction refers to how vertebrae $x \in \{C2, C3, C4\}$ are localized or segmented prior to stage prediction. 'Model-Expert Agreement' indicates whether automated predictions were compared with expert variability. Abbreviations: ROI (Region of Interest), CNN (Convolutional Neural Network), Trf (Transformer).

Work	**Extraction of x**	**Model, Continuous?**	**Public Dataset**	**Model-Expert Agreement**
Kim et al. (2021) [7]	Vertebral mask	CNN, No	No	No
Manoochehri et al. (2022) [8]	ROI + Vertebral mask	CNN, No	No	No
Atici et al. (2022) [9]	Vertebral edges	Custom CNN, No	Yes, Unannotated	No
Atici et al. (2023) [10]	Vertebral edges	Custom CNN, No	Partial	No
Li et al. (2023) [11]	Vertebral mask	CNN, No	No	Yes
Kayaoğlu et al. (2025) [12,13]	Vertebral mask	(CNN, Trf), No	No	No
Motie et al. (2025) [14]	ROI	CNN, No	No	No
ORACLE	**Vertebral mask**	**CNN, Yes**	**Yes**	**Yes**

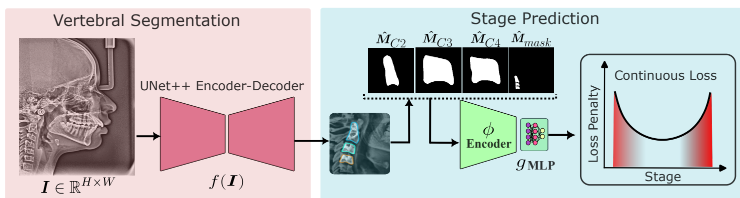

Fig. 1. Overview of the ORACLE pipeline. A LC radiograph is first processed by a `U-Net++` encoder–decoder with an `EfficientNet-B5` encoder to segment cervical vertebrae, from which vertebra-specific masks are extracted. These masks are encoded and passed to an MLP trained with CORAL ordinal loss for continuous stage prediction.

2 Methodology

Let $\boldsymbol{I} \in \mathbb{R}^{H\times W}$ denote a LC radiograph, where H and W represent image height and width, respectively. The objective is to automatically estimate the CVM stage for $\boldsymbol{I}$: i.e. we wish to learn the mapping:

$$m_\theta : \boldsymbol{I} \to \{z\}_{z=1}^{K},$$

where $z \in \{1, \dots, K\}$ denotes the predicted CVM stage and θ denotes the parameters of the learnt model. An overview of the complete pipeline is illustrated in Fig. 1.

2.1 Vertebral Segmentation

A multi-class segmentation model was trained to identify vertebrae C2, C3, and C4. The model predicts a segmentation mask:

$$\hat{\boldsymbol{M}}_{\boldsymbol{mask}} = f(\boldsymbol{I}), \quad \hat{\boldsymbol{M}}_{\boldsymbol{mask}} \in \{0, 1, 2, 3\}^{H\times W},$$

where class 0 denotes background and classes 1–3 correspond to vertebrae C2–C4. CVM staging depends on detecting the boundary of the vertebrae columns precisely. Therefore, it was crucial to give special weight to the boundary pixels during segmentation. The gradients were computed using a composite objective function:

$$\mathscr{L}_{seg} = \lambda_1 \mathscr{L}_{CE} + \lambda_2 \mathscr{L}_{Tversky} + \lambda_3 \mathscr{L}_{Lovasz} + \lambda_4 \mathscr{L}_{Boundary} + \lambda_5 \mathscr{L}_{FocalBoundary}$$

where $\mathscr{L}_{CE}$ is the cross-entropy loss for pixel-wise supervision, $\mathscr{L}_{Tversky}$ is the Tversky loss to mitigate class imbalance, $\mathscr{L}_{Lovasz}$ is the Lovász loss as a surrogate for intersection-over-union optimization, and the boundary terms encourage accurate contour localization. The weighting coefficients $\lambda_{1\ldots5}$ are treated as learnable parameters optimized dynamically during training, allowing the model to automatically balance the relative contribution of each loss component based on its internal uncertainty [20].

2.2 Stage Prediction Model

To reflect the inherent continuity of cervical vertebral maturation, we study three complementary learning formulations: (i) ordinal classification using EMD Loss, (ii) ordinal classification using CORAL Loss, and (iii) continuous regression framework. All approaches share an identical feature extraction pipeline and differ only in the prediction head and loss formulation.

Specifically, to determine the CVM stage, we follow the clinical practice of assessing the shape and concavity of the C2, C3, and C4 vertebrae. Accordingly, we extract cropped segmentation masks for each vertebra ($\hat{\boldsymbol{M}}_{C2}$, $\hat{\boldsymbol{M}}_{C3}$, $\hat{\boldsymbol{M}}_{C4}$), along with the overall cervical mask $\hat{\boldsymbol{M}}_{mask}$. Each mask is independently encoded using a shared image encoder $\phi(\cdot)$. The embeddings are then concatenated to form a structured feature representation $\boldsymbol{h}$ that is passed to a task-specific prediction head $g(\cdot)$.

Regression-Based Continuous Stage Prediction: In the regression formulation, CVM stage prediction is modeled as a continuous estimation problem. The prediction head outputs a single scalar $\hat{y} = g(\boldsymbol{h}) \in \mathbb{R}$. Ground-truth CVM stages are mapped to integer targets in $\{1, \ldots, K\}$. The model is optimized using the mean squared error (MSE) loss.

Ordinal Classification using EMD Loss: The prediction head outputs K class logits, which are converted into a probability distribution $\boldsymbol{p} \in \mathbb{R}^K$ using the softmax function. The ground-truth label is encoded as a one-hot vector $\boldsymbol{y}_{\text{one-hot}}$. To incorporate ordinal structure, we adopt an EMD loss computed as the squared ℓ_2 distance between the Cumulative Distribution Functions (CDFs) of the predicted and target distributions:

$$\mathscr{L}_{\text{EMD}} = \|\text{CDF}(\boldsymbol{p}) - \text{CDF}(\boldsymbol{y}_{\text{one-hot}})\|_2^2$$

Ordinal Classification using CORAL Loss: In the CORAL formulation, the prediction head outputs $\hat{\boldsymbol{y}} \in \mathbb{R}^{K-1}$, corresponding to ordered threshold logits.

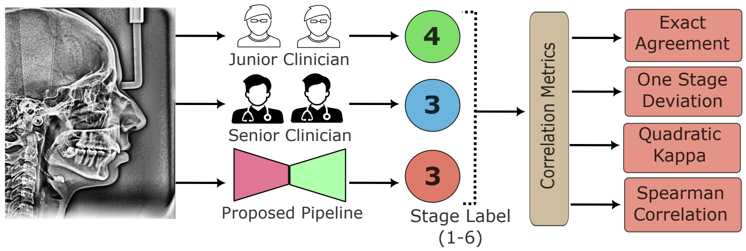

Fig. 2. Overview of the result analysis from the ORACLE pipeline. Each LC radiograph is independently labelled by two junior (≥3 years of experience) and two senior clinicians (≥10 years of experience).

Each output models the probability that the true stage exceeds a given ordinal threshold. The CORAL loss is defined as:

$$\mathscr{L}_{\text{CORAL}} = -\sum_{k=1}^{K-1} \left(\mathbb{I}_{y>k} \log \sigma(\hat{y}) + \mathbb{I}_{y\leq k} \log\left(1-\sigma(\hat{y})\right)\right),$$

where $\sigma(\cdot)$ denotes the sigmoid function and $\mathbb{I}_{(\cdot)}$ is the indicator function. Class probabilities are computed from consecutive differences of the cumulative threshold probabilities.

2.3 Performance Evaluation

There is inter-expert variability in determining the CVM stages, which is invariably reflected in our training set. To understand how ORACLE's predictions relate to predictions from multiple blind experts, we compared its model predictions with the predictions from a set of four clinicians on a separate test set. Two junior clinicians (≥3 years experience) and two senior clinicians (≥10 years experience) were recruited. We measured the agreement of the predictions among the clinicians and with the model predictions. An overview of the result analysis pipeline is illustrated in Fig. 2. Beyond exact accuracy, tolerance-based metrics were reported. Tol-1 and Tol-2 represent the fraction of predictions whose stage differs from the ground-truth stage by at most 1 and 2 stages, respectively.

3 Dataset

This study presents a retrospective cohort of 1,248 LCs sourced from a single high-volume orthodontic center (2019-2024), capturing real-world clinical diversity across pediatric/adolescent patients undergoing routine skeletal maturity assessment. Each LC received dual-expert CVM staging by two senior orthodontists (≥10 years experience) following Baccetti criteria, with adjudication protocol resolving initial disagreements (9% of the cases). The subjects span ages 8-16 years, mixed gender, diverse malocclusions, 205, 163, 173, 277, 244 and 186 LCs for stages 1 to 6, reflecting representative orthodontic caseload. A stratified subset of 569 LCs (46% of cohort) was polygon-annotated for C2/C3/C4 vertebrae using Roboflow Annotate v2.1, with all 569 annotations reviewed and corrected when necessary by two senior orthodontists, ensuring anatomical precision. An independent 113-LC cohort, collected from distinct patient population and different time period, was blind-labeled by 2 junior (≥3yrs) & 2 senior (≥10yrs). The study was an IEC approved retrospective analysis, HIPAA-compliant de-identification was performed (18 PHI fields removed), with no patient identifiers in the released dataset.

4 Experiments and Results

4.1 Experimental Setup

All models were implemented in Python using the `PyTorch` library and trained on a NVIDIA A6000 GPU (48 GB VRAM). All models were evaluated using a fixed 5-fold cross-validation scheme with CVM stage stratification; all vertebrae from the same subject stay together. Within each fold, the segmentation model was trained exclusively on the manually annotated subset. For stage prediction, connected component filtering was applied to retain anatomically consistent vertebral masks. Segmentation quality was assessed using Foreground Intersection-over-Union (IoU), Foreground Dice Coefficient (Dice), and 95th Percentile Hausdorff Distance (HD95).

Classification performance was evaluated using Exact Accuracy and Macro F1-Score. To reflect clinically acceptable deviations, Tol-1 and Tol-2 stages were also computed. Inter-expert and model-expert agreement were assessed using exact agreement, one-stage tolerance agreement, quadratic weighted kappa, and Spearman rank correlation. Based on consultation with two senior orthodontists involved in this study, a ±1 stage deviation is generally considered clinically acceptable in routine practice.

4.2 Segmentation Results

The segmentation model was implemented using `U-Net++`[21] with an EfficientNet-B5[22] encoder pretrained on ImageNet. Training was conducted for 150 epochs with a batch size of 4, utilizing the `AdamW`[23] optimizer (LR of 3×10^{-4}, weight decay of 10^{-5}), a `Cosine Annealing Warm Restarts`[24] scheduler and standard data augmentations. Quantitative performance averaged across five folds on the annotated subset (569 LCs) yielded a Dice of $0.941_{0.007}$, an IoU of $0.888_{0.012}$, and HD95 of $2.719_{0.098}$. Values are reported as mean_{std}.

4.3 Classification Results

Classification models were trained for up to 150 epochs with a batch size of 16, utilizing the `AdamW`[23] optimizer (LR of 10^{-4}, weight decay of 10^{-3}), a `Cosine`

`Annealing Warm Restarts`[24] scheduler, standard augmentations alongside an early stopping mechanism with a patience of 30 epochs. We first performed CNN backbone ablation, where we looked at the performance for different backbones using CORAL Ordinal Loss (summarized in Table 2). Although the tolerances across models are similar, the `ConvNeXt` backbone gives the highest accuracy and F1 scores (4.3% and 7.2% gain over the second best baseline).

The best-performing backbone was then evaluated across several continuous modeling approaches: EMD loss classification, CORAL ordinal classification, and MSE loss regression, alongside a standard CE loss classification baseline. The results of this comparison are detailed in Table 3. The continuous modeling strategies demonstrated improved performance over the standard CE loss method, suggesting that incorporating ordinal structure may better reflect the gradual progression underlying CVM. Exact accuracy remained moderate, reflecting the inherent difficulty of strict, discrete stage assignment. In contrast, tolerance-based performance was substantially higher, indicating that the majority of classification errors occurred between adjacent stages.

Table 2. Five-fold cross-validated test performance for CVM stage prediction using CORAL Ordinal Loss. The evaluated backbone architectures include `ResNet-152`[25], `EfficientNet-B5`[22], `DenseNet-201`[26], and `ConvNeXt-Base`[27]. Values are reported as mean$_{std}$. The best and second-best results are bold and underlined, respectively.

Metric	**CORAL Loss**			
	`ResNet`	`EfficientNet`	`DenseNet`	`ConvNeXt`
Accuracy ↑	$0.563_{0.035}$	$0.534_{0.048}$	$\underline{0.570_{0.040}}$	$\mathbf{0.628_{0.041}}$
F1 Score ↑	$0.525_{0.054}$	$0.491_{0.063}$	$\underline{0.551_{0.044}}$	$\mathbf{0.622_{0.040}}$
Tol-1 ↑	$\underline{0.944_{0.028}}$	$0.927_{0.022}$	$0.933_{0.024}$	$\mathbf{0.957_{0.021}}$
Tol-2 ↑	$\mathbf{0.996_{0.005}}$	$0.991_{0.007}$	$\underline{0.993_{0.005}}$	$\mathbf{0.996_{0.005}}$

4.4 Inter-Expert Variability

To contextualize automated performance, a separate dataset of 113 LCs, distinct from the training and evaluation data, was independently staged by ORACLE (`ConvNeXt-Base` + CORAL Loss) and two sets of clinicians (independently). Predictions from the five fold-specific models were averaged to obtain the final stage assignment.

As shown in Fig. 3, the exact agreement between the model and the four clinicians ranged from 37.5% to 65.2%, comparing favorably to the strict exact match rate between the two senior experts (36.6%). Furthermore, the model's adjacent-stage agreement ranged from 76.8% to 95.5%, confining most disagreements to neighboring stages and performing on par with the observed inter-clinician range of 58.9% to 97.3%. In terms of reliability, the model achieved κ scores between 0.638 and 0.891, exceeding the inter-senior baseline ($\kappa = 0.738$) against three of

Table 3. Five-fold cross-validated test performance of CVM stage prediction on segmentation-derived binary masks using the `ConvNeXt-Base`. The conventional categorical baseline, Cross-Entropy (CE), is compared against continuous learning formulations (EMD Classification, CORAL Classification, and Regression using MSE).

Metric	**ConvNeXt**			
	CE Loss	EMD Loss	CORAL Loss	MSE Loss
Accuracy ↑	$0.584_{0.023}$	$0.619_{0.041}$	$\mathbf{0.628_{0.041}}$	$0.599_{0.032}$
F1 Score ↑	$0.568_{0.019}$	$0.609_{0.035}$	$\mathbf{0.622_{0.040}}$	$0.590_{0.032}$
Tol-1 ↑	$0.948_{0.019}$	$\mathbf{0.961_{0.020}}$	$0.957_{0.021}$	$\mathbf{0.961_{0.011}}$
Tol-2 ↑	$0.995_{0.005}$	$0.996_{0.004}$	$0.996_{0.006}$	$\mathbf{0.997_{0.005}}$

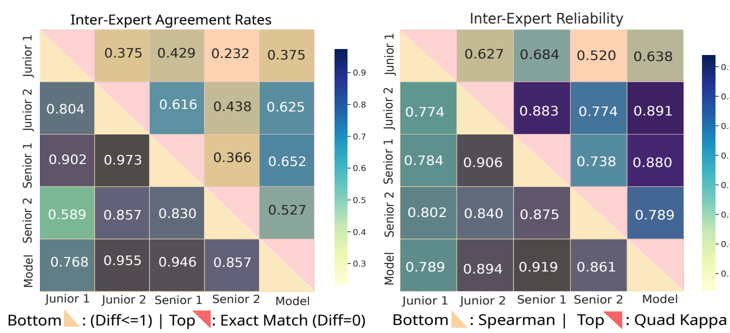

Fig. 3. Metrics for CVM stage classification on the held out dataset. **Left:** Inter-expert agreement among model and clinicians. **Right:** Correlation between the model and experts. Juniors have ≥3 years experience, while Seniors have ≥10 years experience.

the four clinicians. Finally, consistently high Spearman correlations ($\rho \geq 0.789$) validate that the continuous formulation effectively enforces the ordinality required for staging.

5 Conclusion

This study provides evidence that CVM staging can benefit from continuous modeling approaches in addition to conventional categorical classification, as skeletal maturation exhibits gradual transitions poorly captured by discrete boundaries - evidenced by model Tol-1 accuracy 95.7-96.1% vs exact 62.8%, matching clinician inter-expert patterns (junior/senior $\kappa = 0.52 - 0.88$, senior/senior $\kappa = 0.74$). `ConvNeXt-Base` + CORAL ordinal loss outperforms the CE baseline (+3.5% accuracy, +1.3% Tol-1) and all other formulations, indicating that exploiting ordinal structure can improve agreement with clinically accepted evaluation criteria. Expert validation demonstrates model agreement with senior orthodontists ($\kappa = 0.88/0.79$) falls within the range of observed inter-expert variability. We introduce the first comprehensive public dataset with

1,248 LC stage labels including 569 expert-verified C2-C4 polygon masks, addressing computational orthodontics' data scarcity. Future work includes continuous maturity scoring, Bayesian uncertainty for borderline cases, prospective clinical trials measuring treatment timing accuracy, and multi-center validation. Public code and data release establish a reproducible benchmark for precision orthodontics AI.

Disclosure of Interests

The authors have no competing interests to declare.

References

1. T. Baccetti, L. Franchi, and J. A. McNamara, "The cervical vertebral maturation (cvm) method for the assessment of optimal treatment timing in dentofacial orthopedics," *Seminars in Orthodontics*, vol. 11, no. 3, pp. 119–129, 2005.
2. T. Baccetti, L. Franchi, and J. A. McNamara Jr, "An improved version of the cervical vertebral maturation (cvm) method for the assessment of mandibular growth," *The Angle Orthodontist*, vol. 72, no. 4, pp. 316–323, 2002.
3. T. S. Nestman, S. D. Marshall, F. Qian, N. Holton, R. G. Franciscus, and T. E. Southard, "Cervical vertebrae maturation method morphologic criteria: Poor reproducibility," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 140, no. 2, pp. 182–188, 2011.
4. X.-G. Zhao, J. Lin, J.-H. Jiang, Q. Wang, and S. H. Ng, "Validity and reliability of a method for assessment of cervical vertebral maturation," *The Angle Orthodontist*, vol. 82, no. 2, pp. 229–234, 2012.
5. U. Hussain, F. U. Hassan, M. A. Kamran, A. A. Alnazeh, I. Ullah, and S. Alam, "Inter-observer and intra-observer agreement of cervical vertebral maturation staging: A systematic review and meta-analysis," *International Orthodontics*, vol. 22, no. 3, p. 100874, 2024.
6. G. Perinetti, A. Caprioglio, and L. Contardo, "Visual assessment of the cervical vertebral maturation stages: A study of diagnostic accuracy and repeatability," *The Angle Orthodontist*, vol. 84, no. 6, pp. 951–956, 2014.
7. E.-G. Kim, I.-S. Oh, J.-E. So, J. Kang, V. N. T. Le, M.-K. Tak, and D.-W. Lee, "Estimating cervical vertebral maturation with a lateral cephalogram using the convolutional neural network," *Journal of Clinical Medicine*, vol. 10, no. 22, 2021.
8. H. Manoochehri, S. A. Motamedi, A. Mohammad-Djafari, M. Makaremi, and A. Vafaie Sadr, "Attention-guided multi-scale cnn network for cervical vertebral maturation assessment from lateral cephalometric radiography," in *Physical Sciences Forum*, vol. 5, p. 26, MDPI, 2022.
9. S. F. Atici, R. Ansari, V. Allareddy, O. Suhaym, A. E. Cetin, and M. H. Elnagar, "Fully automated determination of the cervical vertebrae maturation stages using deep learning with directional filters," *Plos one*, vol. 17, no. 7, p. e0269198, 2022.
10. S. F. Atici, R. Ansari, V. Allareddy, O. Suhaym, A. E. Cetin, and M. H. Elnagar, "Aggregatenet: A deep learning model for automated classification of cervical vertebrae maturation stages," *Orthodontics & craniofacial research*, vol. 26, pp. 111–117, 2023.

11. H. Li, H. Li, L. Yuan, C. Liu, S. Xiao, Z. Liu, G. Zhou, T. Dong, N. Ouyang, L. Liu, *et al.*, "The psc-cvm assessment system: A three-stage type system for cvm assessment based on deep learning," *BMC Oral Health*, vol. 23, no. 1, p. 557, 2023.
12. M. Kayaoğlu, A. Şengür, S. Ç. Ciğerim, and S. Bor, "Detection of cervical vertebrae using object detection and semantic segmentation methods in lateral cephalometric radiographs," *Türk Doğa ve Fen Dergisi*, vol. 14, no. 2, pp. 26–36, 2025.
13. M. Kayaoğlu, A. Şengür, S. Bor, and S. Kotan, "Classification of cervical vertebral maturation stages and bone age assessment using transfer learning–based deep-learning approaches," *Firat University Journal of Experimental and Computational Engineering*, vol. 4, no. 2, pp. 393–405, 2025.
14. P. Motie, A. Ashkan, H. Mohammad-Rahimi, S. Hassanzadeh-Samani, N. Razzaghi, M. Behnaz, S. Shahab, and S. R. Motamadian, "Improving cervical maturation degree classification accuracy using a multi-stage deep learning approach," *Imaging Science in Dentistry*, vol. 55, no. 3, p. 290, 2025.
15. W. Cao, V. Mirjalili, and S. Raschka, "Rank consistent ordinal regression for neural networks with application to age estimation," *Pattern Recognition Letters*, vol. 140, pp. 325–331, 2020.
16. L. Hou, C.-P. Yu, and D. Samaras, "Squared earth mover's distance-based loss for training deep neural networks," *arXiv preprint arXiv:1611.05916*, 2016.
17. Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal regression with multiple output cnn for age estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4920–4928, 2016.
18. X. Liu, Y. Zou, Y. Song, C. Yang, J. You, and B. K Vijaya Kumar, "Ordinal regression with neuron stick-breaking for medical diagnosis," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pp. 0–0, 2018.
19. P. Chen, L. Gao, X. Shi, K. Allen, and L. Yang, "Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 84–92, 2019.
20. A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7482–7491, 2018.
21. Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *International workshop on deep learning in medical image analysis*, pp. 3–11, Springer, 2018.
22. M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, pp. 6105–6114, PMLR, 2019.
23. I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.
24. I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.
25. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
26. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.
27. Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11976–11986, 2022.