Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Artificial Intelligence for Tooth Numbering and Full-Mouth Series Placement

Britney Nguyen
The University of Texas Health Science Center at Houston School of Dentistry
Kearny Chang
The University of Texas Health Science Center at Houston School of Dentistry
Yi Liu
National Yang Ming Chiao Tung University
Nikola Angelov
The University of Texas Health Science Center at Houston School of Dentistry
Ming-Feng Chang
National Yang Ming Chiao Tung University
Aaron Glick
The University of Texas Health Science Center at Houston School of Dentistry
Hsiu-Wan Meng
The University of Texas Health Science Center at Houston School of Dentistry
Celine Sheng
The University of Texas Health Science Center at Houston School of Dentistry
Sally Sheng
The University of Texas Health Science Center at Houston School of Dentistry
Srinivas Ayilavarapu
The University of Texas Health Science Center at Houston School of Dentistry
Jennifer Chang
jennifer.chang@uth.tmc.edu

The University of Texas Health Science Center at Houston School of Dentistry

Article

Keywords: tooth detection, object detection, deep learning, Siamese neural networks, convolutional neural networks, dental radiographs

Posted Date: October 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10974409/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

This study proposes an automated method for detecting tooth positions and predicting full-mouth series (FMS) radiograph placement using oriented object detection and Siamese neural networks. A total of 5,431 radiographs from 354 patients were annotated and used to train deep learning models. Oriented bounding boxes were generated using a YOLOv5 detector, followed by tooth numbering, feature extraction, and image classification via a multitask ResNet50 model. FMS placement was predicted using a Siamese neural network inspired by jigsaw-puzzle tasks, with permutation-based training and patient-level five-fold cross-validation. Mean tooth-numbering accuracy was 0.99 ± 0.00. Mean FMS-placement accuracy was 0.97 ± 0.01 for upper left and right periapical radiographs, 0.98 ± 0.02 for upper anterior periapical radiographs, 0.98 ± 0.01 for lower left and right periapical radiographs, 0.99 ± 0.01 for lower anterior periapical radiographs, and 0.96 ± 0.02 for bitewing radiographs; overall mean accuracy was 0.98 ± 0.01. This internally tested AI pipeline demonstrated high accuracy for tooth numbering and automated FMS arrangement and may help streamline dental radiographic workflows, subject to external and clinical validation.

Introduction

Artificial Intelligence (AI), first proposed in the 1950s, remained underutilized for decades due to limited computational resources. Recent advancements in processing power and data storage have accelerated its development, particularly in machine learning and deep learning (DL). DL models excel in processing large datasets, including medical images, and have shown strong performance in diagnostics [1–3], financial forecasting [4, 5], and autonomous driving [6, 7].

In medical imaging, DL models can automatically detect, classify, and label anatomical structures and abnormalities with high accuracy [8, 3]. Convolutional Neural Networks (CNNs), in particular, are widely used for such tasks. CNNs are supervised DL algorithms inspired by the organization and function of the mammalian visual system [9]. Through convolutional and pooling layers, these models identify key image features while reducing dimensionality and computational complexity. Extracted features are then processed through nonlinear transformations and fully connected layers to produce the final output.

In dental imaging, prior DL studies demonstrated the feasibility of automated tooth detection and numbering on periapical and bitewing radiographs [10–13]. Subsequent studies developed an AI framework for tooth-number recognition in panoramic and intraoral radiographs and automated arrangement of intraoral radiographs into a full-mouth series (FMS) X-ray template [14]. However, image placement based on tooth numbering alone is inherently vulnerable to incomplete or ambiguous radiographic anatomy and is often unreliable for bitewing images. This reduced anatomic context combined with suboptimal plate positioning and overlapping contacts can make reliable tooth numbering difficult and may require manual correction [15]. To address these limitations, this study aims to evaluate a more generalized and automated placement strategy. An end-to-end AI pipeline for automated tooth detection, numbering, and image-level spatial prediction for FMS placement.

The proposed framework leverages YOLOv5 oriented bounding box (OBB) detection for spatially precise tooth localization, a multitask ResNet50 model for simultaneous tooth detection and numbering, and a Siamese neural network for predicting the spatial arrangement of periapical and bitewing radiographs within a standardized FMS template [16–18]. Inspired by jigsaw puzzle logic, this framework reduces dependence on tooth-level labels and supports automated bitewing placement by performing FMS placement with both an object-recognition and spatial-relationship approach.

Materials and Methods

Data Collection

This retrospective study was approved by the Committee for the Protection of Human Subjects of The University of Texas Health Science Center at Houston (HSC-DB-19-0994). All methods were performed in accordance with the relevant guidelines and regulations, including the principles of the Declaration of Helsinki. The requirement for informed consent was waived by the institutional review board because the study used previously collected, de-identified radiographic data.

A total of 5,431 FMS radiographs from 354 patients were used for the dataset (Table 1). Collected periapical and bitewing radiographs with personal information were de-identified and stored in the encrypted cloud drive. Each radiograph was reviewed and annotated by dental professionals with corresponding universal tooth numbers and the appropriate FMS placement, establishing the ground-truth against which the model predictions were evaluated for accuracy. Annotating was conducted using the Labelme annotation tool [19], with bounding boxes drawn for each of the 6,085 teeth.

Table 1
Performance of the tooth numbering and FMS placement models

Tooth Numbering						
	Fold 1	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Avg ± SD**
Accuracy	0.99	1.00	0.99	0.99	1.00	0.99 ± 0.00
FMS Placement						
Region	**Fold 1**	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Avg ± SD**
ULR	0.98	0.96	0.97	0.99	0.96	0.97 ± 0.01
UMD	0.96	0.98	0.99	0.99	0.96	0.98 ± 0.02
LLR	0.97	0.98	0.99	0.98	0.99	0.98 ± 0.01
LMD	0.99	1.00	0.99	0.99	1.00	0.99 ± 0.01
BW	0.97	0.98	0.95	0.94	0.98	0.96 ± 0.02
Overall	0.97	0.98	0.98	0.98	0.98	0.98 ± 0.01

Abbreviations: ULR, upper left and right periapical radiographs; UMD, upper middle periapical radiographs; LLR, lower left and right periapical radiographs; LMD, lower middle periapical radiographs; BW, bitewing radiographs.

Neural Network Architecture

Three deep learning models were incorporated into an automated radiographic analysis framework: YOLOv5-OBB for tooth detection, ResNet50-based CNNs for tooth numbering, feature extraction, and image classification, and a Siamese neural network for FMS placement. An overview of the study workflow is presented in Fig. 1.

YOLOv5-OBB was used to detect individual teeth within each radiograph while accounting for variations in tooth angulation and image orientation. Each detected tooth region was then processed using ResNet50-based CNN to extract high-level and relevant visual and spatial characteristics to predict tooth numbers. The final tooth-numbering results were compared with the ground-truth, and accuracy was calculated to evaluate the performance and reliability of the entire processing workflow.

For FMS placement, each augmented input radiograph was processed by a supervised three-class ResNet50 model, which extracted anatomical regions and image aspect ratios before classifying the radiograph as a maxillary periapical, mandibular periapical, or bitewing radiograph. Maxillary and mandibular periapical radiographs were further categorized into the upper left and right (ULR), upper middle (UMD), lower left and right (LLR), and lower middle (LMD) regions, while bitewing radiographs formed a separate bitewing (BW) group.

The Siamese network architecture used for each region was determined by the number of radiographs in that group. A 3-input/6-class network was used for the LMD region, which contained three radiographs. Separate 4-input/24-class networks were used for the ULR, LLR, and BW regions, each of which contained four radiographs. A 5-input/120-class network was used for the UMD region, which contained five radiographs. The output classes represented all possible image-order permutations within each region: 3! = 6, 4! = 24, and 5! = 120. By extracting and comparing features across each group, the networks learned the relative spatial relationships among the radiographs and predicted their anatomical order. These positional predictions, together with the initial image classifications, were then used to final FMS organization.

Figure 2 provides a visual overview of the anatomical regions, image groupings, and corresponding network inputs. The ULR region included positions 1, 2, 8, and 9; the UMD region included positions 3 through 7; the LLR region included positions 10, 11, 15, and 16; the LMD region included positions 12 through 14; and the BW region included positions 17 through 20.

Training Strategy

To ensure generalization of the constructed CNN model, five-fold cross-validation was implemented. The original dataset was randomly divided into five equal-sized folds. In each iteration, three folds were used for training, one fold was used for validation, and one fold was reserved for testing. The assignments were rotated across the five iterations so that each fold served once as the validation set and once as the testing set. The training data were used to develop the CNN model, the validation data were used for model selection, and the testing data were used to evaluate the performance of the selected model. Performance metrics were then summarized across the five test folds. Data leakage was avoided by ensuring all radiographs and teeth from the same patient were assigned to a single group and not split across the training, validation, and testing folds. Doing so prevented the model from being evaluated on data closely resembling the data used during training, which could lead to an overestimation of its accuracy.

Data augmentation procedures included horizontal flipping, contrast adjustment (0–30%), brightness variation (0–50%), and rotation (± 5°). Because the training set was created by randomly sampling from the full augmented dataset, not every original tooth image was guaranteed to be represented in the selected samples; however, the likelihood of an original image being omitted was low. During testing, only non-augmented data were used so that the model's predictive performance would reflect real-world conditions.

Model training was initially performed using stochastic gradient descent, an optimization method that adjusts the model's internal weights step by step to reduce errors. Settings were a learning rate of 1×10^{-5}, batch size of 32, and cross entropy loss. In later training stages, AdamW was used as the optimizer to fine-tune the model by adaptively updating the model weights while applying decoupled weight decay to improve training stability and reduce overfitting [20].

To predict missing teeth, a shallow neural network was used [10]. The model took horizontal distances between adjacent bounding boxes as input and predicted the presence of 0, 1, or 2 missing teeth. The architecture included two hidden layers with dropout (0.2) to reduce overfitting, and training was performed over fifty epochs using SGD and cross entropy loss.

Tooth numbering followed the Universal Tooth Numbering system, in which the permanent maxillary teeth are numbered 1–16 from right to left and the permanent mandibular teeth are numbered 17–32 from left to right [21].

For tooth numbering, intersection over union (IoU) was calculated to assess the degree of overlap between each predicted bounding box and the corresponding ground-truth bounding box. Predictions with an IoU value greater than 0.70 were considered successful matches, and the prediction with the highest IoU for each ground-truth box was selected as the best match. The predicted and ground-truth tooth numbers from the matched bounding-box pairs were then compared to calculate tooth-numbering accuracy.

For FMS placement, each radiograph was classified directly into its corresponding FMS position; therefore, IoU-based matching was not required, and accuracy was calculated by comparing the predicted position with the ground-truth position.

Results

Using an IoU threshold greater than 0.70, tooth-numbering accuracy was 0.99 for Fold 1, 1.00 for Fold 2, 0.99 for Fold 3, 0.99 for Fold 4, and 1.00 for Fold 5, yielding a mean accuracy of 0.99 ± 0.00 across the five cross-validation folds (Table 1). The predicted and ground-truth tooth numbers from the matched bounding-box pairs were then compared to calculate tooth-numbering accuracy, with the results summarized in a confusion matrix (Fig. 3). Most values are concentrated along the main diagonal, indicating that the predicted positions closely matched the reference positions and supporting the reliability of the proposed method.

The combined confusion matrices were generated by pooling out-of-fold predictions from all five cross-validation folds. The pooled analysis included 1,380 upper left and right periapical observations, 965 upper middle periapical observations, 1,388 lower left and right periapical observations, 1,062 lower middle periapical observations, and 636 bitewing observations, for a total of 5,431 predictions. Of these, 5,316 were correctly classified, corresponding to a pooled accuracy of 0.98. The matrices demonstrated strong diagonal dominance, indicating high agreement between the ground-truth and predicted FMS positions. Misclassifications were uncommon and were primarily confined to closely related or adjacent classes rather than broadly distributed across unrelated positions.

Accuracy of FMS placement for positions 1 through 20 is represented in a confusion matrix (Fig. 4). The confusion matrix demonstrated strong placement performance across all twenty positions, with a mean position-specific correct placement rate of 0.98 and diagonal values ranging from 0.96 to 1.00. Position

13 achieved perfect placement, while positions 5, 12, and 14 each reached approximately 0.99. The lowest correct placement rates were observed for positions 6, 7, 19, and 20, although each remained high at 0.96. Incorrect placements were infrequent and were primarily concentrated between neighboring or closely related position pairs, including positions 1 and 2, 3 and 4, 6 and 7, 8 and 9, 10 and 11, 15 and 16, 17 and 18, and 19 and 20. The largest off-diagonal error was 0.04 between positions 19 and 20. Overall, the model demonstrated high accuracy in assigning images to the correct positions, with only limited confusion between closely associated positions.

Using five-fold cross-validation, FMS placement accuracy for the ULR region was 0.98, 0.96, 0.97, 0.99, and 0.96 across Folds 1 through 5, respectively, yielding a mean accuracy of 0.97 ± 0.01. For the UMD region, accuracy was 0.96, 0.98, 0.99, 0.99, and 0.96, with a mean of 0.98 ± 0.02. For the LLR region, accuracy was 0.97, 0.98, 0.99, 0.98, and 0.99, with a mean of 0.98 ± 0.01. For the LMD region, accuracy was 0.99, 1.00, 0.99, 0.99, and 1.00, with a mean of 0.99 ± 0.01. Bitewing placement accuracy was 0.97, 0.98, 0.95, 0.94, and 0.98, yielding a mean of 0.96 ± 0.02. Overall FMS placement accuracy across the five folds was 0.97, 0.98, 0.98, 0.98, and 0.98, with a mean accuracy of 0.98 ± 0.01 (Table 1).

Discussion

For tooth detection, YOLOv5-based oriented object detectors using OBBs were employed to accurately localize teeth across varying rotations while maintaining real-time detection efficiency [22]. In intraoral radiography, tooth orientation and projection geometry can vary depending on sensor placement, patient positioning, and radiographic angulation. Compared to conventional horizontal bounding boxes, OBB detection is better suited for localizing objects that appear tilted, rotated, partially cropped, or projected at varying angles [23]. Candidate models were screened based on detection performance and computational cost, with Oriented R-CNN used as a benchmark given its established two-stage approach to oriented object detection [24]. Its first stage generates oriented region proposals, and its second stage classifies and refines those proposals to produce final rotated bounding boxes. This benchmarking framework enabled selection of the detector that provided the best balance between localization accuracy and computational efficiency for subsequent automated FMS placement.

Previous studies using periapical radiographs reported automated tooth-numbering accuracies of 0.87 with Faster R-CNN and Inception v3 and 0.91 with Faster R-CNN and rule-based postprocessing [11, 10]. Studies using panoramic radiographs reported accuracies of 0.88 with YOLOv4, 0.98 with Faster R-CNN and VGG-16, and 0.99 with an optimized YOLOv7 model [25, 26, 21]. Use of YOLOv5 OBB in this study led to a substantial increase in detection speed while maintaining a high average accuracy of 0.99. High precision is particularly important in this setting, as false positives may distort the detected tooth count and spatial configuration, thereby propagate errors into downstream tooth numbering and FMS placement. In this study, false positives were primarily due to unannotated teeth at the edges of the image, and false negatives were associated with incomplete tooth labeling, overlapping molars, extreme tooth rotations, and structural anomalies such as fractures. The high average accuracy of 0.99 demonstrates the proposed framework's capability for fast, accurate, and fully automated analysis of

periapical and bitewing radiographs, offering notable advantages over manual and previous automated methods in both performance and clinical relevance.

This study demonstrates notable advancements in automated dental image analysis by reducing reliance on explicit tooth-number recognition for FMS placement. Tooth-number-based methods may be less reliable when radiographs contain incomplete or ambiguous radiographic anatomy, overlapping structures, or missing teeth. This limitation is particularly relevant for bitewing radiographs which primarily capture the crowns and alveolar crests, with substantially less root visualization than periapical radiographs. By combining rule-based bounding box detection with an orientation-aware object detection framework, the proposed methods used spatial distribution and orientation of anatomy to support accurate localization despite variations in tooth position and angulation. This approach reduced reliance on complete tooth morphology and enabled consistent placement across both periapical and bitewing regions.

ResNet50 was selected instead of earlier convolutional architectures (e.g. AlexNet and GoogLeNet) because residual connections facilitate training of deeper networks and have made ResNet architectures widely used as feature-extraction backbones in medical image analysis [17, 27]. ResNet-based multitask models have been used for simultaneous classification of dental implant brand and treatment stage on panoramic radiographs, supporting the feasibility of this approach in dental image analysis [28]. The multitask ResNet50 component enhances the robustness of the overall framework by allowing tooth classification and numbering to be learned jointly. A multitask approach can encourage the network to learn shared visual features including crown shape, root pattern, tooth position, and arch-specific morphology. This is particularly useful for FMS placement, where the same image may contain multiple teeth, only partial anatomy, or ambiguous structures. Multitask learning also reflects clinical reasoning, as dentists do not interpret radiographs using tooth numbers alone, but integrate tooth morphology, location, neighboring anatomy, and restorations. Accordingly, ResNet50 was used here as the classification backbone and shared Siamese feature extractor.

For FMS placement, PubMed searches identified limited prior research specifically evaluating automated FMS arrangements. Kabir et al. used an automated tooth-number-guided framework to arrange periapical and bitewing radiographs. Their method segmented individual teeth, matched them with a panoramic radiograph or tooth repository, and assigned tooth numbers before determining radiograph placement. The study achieved placement accuracies of 0.95 for periapical radiographs and 0.90 for bitewing radiographs. In comparison, the current study used a Siamese neural network to classify radiograph placement directly and achieved accuracies of 0.99 and 0.97, respectively. For this study, a Siamese neural network inspired by the Context-Free Network was developed [29]. The original Context-Free Network used a jigsaw puzzle-solving task to learn spatial relationships between image patches. The Siamese neural network adds a second layer of reasoning by learning whether two radiographs are spatially or anatomically related within the FMS layout. This is where the “jigsaw puzzle” concept becomes especially relevant. In an FMS, each radiograph functions like one piece of a larger spatial arrangement. The goal is not only to identify what teeth are present, but also to determine where each

image belongs relative to the others. A Siamese architecture is well suited for this problem because it compares pairs of images through shared-weight feature extractors and learns similarity or positional relationships between them. In medical imaging, Siamese and self-supervised similarity-learning methods have been used to compare related images, learn meaningful representations, and reduce dependence on exhaustive manual annotation. In dental radiography specifically, recent work has explored Siamese networks for comparing longitudinal dental X-rays, supporting the relevance of this architecture for dental image-pair analysis. More broadly, Siamese architectures compare paired inputs using shared-weight feature extractors and learn similarity or relational features between images [18]. This concept was adapted to assign periapical and bitewing radiographs to their appropriate positions within an FMS template, allowing placement to incorporate spatial relationships rather than relying solely on tooth-level labels. As a result, the need for intermediate tooth labeling is eliminated, reducing error propagation and enabling automated bitewing positioning—an area previously unresolved [14, 10, 26].

The proposed pipeline achieved a mean FMS placement accuracy of 0.98 ± 0.01, including an accuracy of 0.96 ± 0.02 for bitewing placement. These outcomes indicate strong generalizability and clinical applicability. By incorporating OBBs, the method more tightly localized rotated structures and reduced the background inclusion and bounding-box overlap associated with conventional horizontal bounding boxes. Doing so supported improved detection in complex settings involving densely arranged or partially occluded objects [30]. By integrating a ResNet50 backbone, the Siamese framework extracted hierarchical radiographic features, while residual connections supported stable training and preserved spatial consistency and anatomical symmetry, thereby enabling accurate and efficient FMS sequencing.

Conclusion

In summary, this study developed an automated AI-based framework that reduces reliance on manual labeling, addresses limitations of prior approaches, and successfully enables automated bitewing image placement. With its enhanced speed, accuracy, and adaptability, the framework represents a significant advancement in AI-driven dental radiographic analysis and holds promise for improving diagnostic workflows in clinical practice. Future studies should validate the system across larger and more diverse clinical datasets, evaluate its performance in real-world clinical practice, and assess its integration into diagnostic workflows to further enhance its clinical applicability. In addition, future project improvements may include adding panoramic radiographs to improve accuracy of both tooth numbering and FMS placement.

Declarations

Author Contributions

All authors gave final approval and agreed to be accountable for all aspects of the work.

Declaration of Conflicting Interests

The authors declare no conflicts of interest related to the research, authorship, or publication of this article.

Funding

No financial support was received for the research, authorship, or publication of this article.

Data availability

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Oh, S. L. et al. A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Comput. Appl.* **32**, 10927–10933. https://doi.org/10.1007/s00521-018-3689-5 (2020).
2. Ortiz, A., Munilla, J., Górriz, J. M. & Ramírez, J. Ensembles of deep learning architectures for the early diagnosis of Alzheimer's disease. *Int. J. Neural Syst.* **26**, 1650025. https://doi.org/10.1142/S0129065716500258 (2016).
3. Shen, D., Wu, G. & Suk, H. I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **19**, 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442 (2017).
4. Barboza, F., Kimura, H. & Altman, E. Machine learning models and bankruptcy prediction. *Expert Syst. Appl.* **83**, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006 (2017).
5. Sezer, O. B., Gudelek, M. U. & Ozbayoglu, A. M. Financial time series forecasting with deep learning: a systematic literature review: 2005–2019. *Appl. Soft Comput.* **90**, 106181. https://doi.org/10.1016/j.asoc.2020.106181 (2020).
6. Aradi, S. Survey of deep reinforcement learning for motion planning of autonomous vehicles. *IEEE Trans. Intell. Transp. Syst.* **23**, 740–759. https://doi.org/10.1109/TITS.2020.3024655 (2022).
7. Grigorescu, S., Trasnea, B., Cocias, T. & Macesanu, G. A survey of deep learning techniques for autonomous driving. *J. Field Robot*. **37**, 362–386. https://doi.org/10.1002/rob.21918 (2020).
8. Ronneberger, O., Fischer, P. & Brox, T. U-net: convolutional networks for biomedical image segmentation. In *Medical Image Comput. Computer-Assisted Intervention* **9351**, 234–241 (Springer, 2015). https://doi.org/10.1007/978-3-319-24574-4_28
9. Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* **195**, 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455 (1968).
10. Chen, H. et al. A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. *Sci. Rep.* **9**, 3840. https://doi.org/10.1038/s41598-019-40414-y (2019).
11. Görürgöz, C. et al. Performance of a convolutional neural network algorithm for tooth detection and numbering on periapical radiographs. *Dentomaxillofac Radiol.* **51**, 20210246.

https://doi.org/10.1259/dmfr.20210246 (2022).

12. Ayhan, B., Ayan, E. & Bayraktar, Y. A novel deep learning-based perspective for tooth numbering and caries detection. *Clin. Oral Investig*. **28**, 178. https://doi.org/10.1007/s00784-024-05566-w (2024).
13. Sadr, S. et al. Deep learning for tooth identification and numbering on dental radiography: a systematic review and meta-analysis. *Dentomaxillofac Radiol.* **53**, 191–203. https://doi.org/10.1093/dmfr/twad003 (2024).
14. Kabir, T., Lee, C. T., Chen, L., Jiang, X. & Shams, S. A comprehensive artificial intelligence framework for dental diagnosis and charting. *BMC Oral Health*. **22**, 480. https://doi.org/10.1186/s12903-022-02514-6 (2022).
15. Cejudo, J. E., Chaurasia, A., Feldberg, B., Krois, J. & Schwendicke, F. Classification of dental radiographs using deep learning. *J. Clin. Med.* **10**, 1496. https://doi.org/10.3390/jcm10071496 (2021).
16. Jocher, G. et al. ultralytics/yolov5: v7.0—YOLOv5 SOTA realtime instance segmentation. *Zenodo* https://doi.org/10.5281/zenodo.7347926 (2022).
17. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778 (IEEE, 2016). https://doi.org/10.1109/CVPR.2016.90
18. Koch, G., Zemel, R. & Salakhutdinov, R. Siamese neural networks for one-shot image recognition. In *ICML 2015 Deep Learning Workshop* (2015).
19. Russell, B. C., Torralba, A., Murphy, K. P. & Freeman, W. T. LabelMe: a database and web-based tool for image annotation. *Int. J. Comput. Vis.* **77**, 157–173. https://doi.org/10.1007/s11263-007-0090-8 (2008).
20. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. In *International Conference on Learning Representations* (2019). https://doi.org/10.48550/arXiv.1711.05101
21. Ali, M. A., Fujita, D. & Kobashi, S. Teeth and prostheses detection in dental panoramic X-rays using CNN-based object detector and a priori knowledge-based algorithm. *Sci. Rep.* **13**, 16542. https://doi.org/10.1038/s41598-023-43591-z (2023).
22. Li, X., Cai, Z. & Zhao, X. Oriented-YOLOv5: a real-time oriented detector based on YOLOv5. In *Proceedings of the 7th International Conference on Computer and Communication Systems*, 216–222 (IEEE, 2022). https://doi.org/10.1109/ICCCS55155.2022.9846234
23. Wen, L., Cheng, Y., Fang, Y. & Li, X. A comprehensive survey of oriented object detection in remote sensing images. *Expert Syst. Appl.* **224**, 119960. https://doi.org/10.1016/j.eswa.2023.119960 (2023).
24. Xie, X., Cheng, G., Wang, J., Yao, X. & Han, J. Oriented R-CNN for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3500–3509 (IEEE, 2021). https://doi.org/10.1109/ICCV48922.2021.00350
25. Putra, R. H. et al. Automated permanent tooth detection and numbering on panoramic radiograph using a deep learning approach. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **137**, 537–544.

https://doi.org/10.1016/j.oooo.2023.06.003 (2024).

26. Tuzoff, D. V. et al. Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dentomaxillofac Radiol.* **48**, 20180051. https://doi.org/10.1259/dmfr.20180051 (2019).
27. Yu, H., Yang, L. T., Zhang, Q., Armstrong, D. & Deen, M. J. Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives. *Neurocomputing* **444**, 92–110. https://doi.org/10.1016/j.neucom.2020.04.157 (2021).
28. Sukegawa, S. et al. Multi-task deep learning model for classification of dental implant brand and treatment stage using dental panoramic radiograph images. *Biomolecules* **11**, 815. https://doi.org/10.3390/biom11060815 (2021).
29. Noroozi, M. & Favaro, P. Unsupervised learning of visual representations by solving jigsaw puzzles. In *European Conference on Computer Vision*, 69–84 (Springer, 2016). https://doi.org/10.1007/978-3-319-46466-4_5
30. Chen, Z. et al. PIoU loss: towards accurate oriented object detection in complex environments. In *Computer Vision—ECCV 2020*, 195–211 (Springer, 2020). https://doi.org/10.1007/978-3-030-58558-7_12

Figures

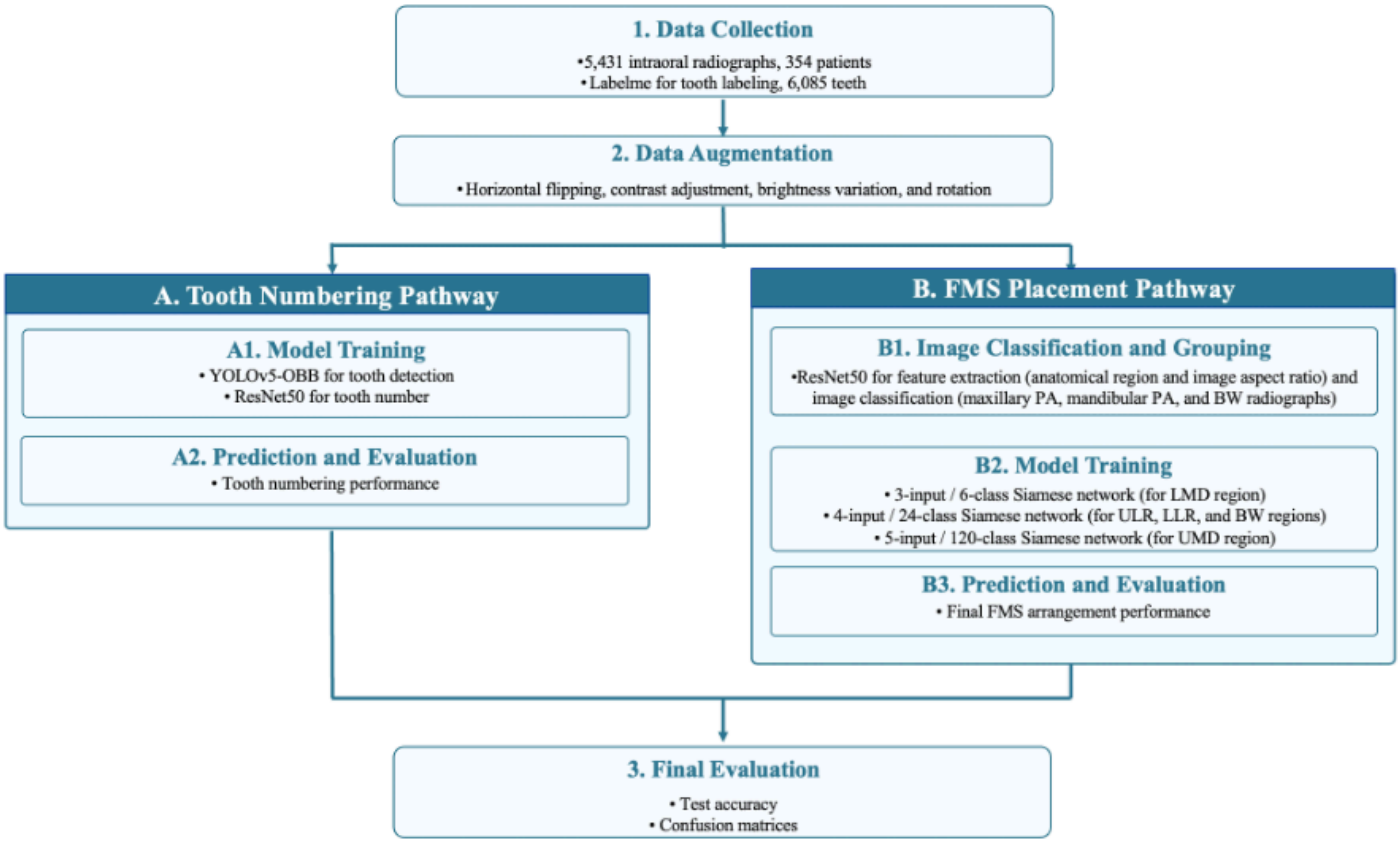


Figure 1

Overview of the study workflow

Abbreviations: PA, periapical radiographs; BW, bitewing radiographs; LMD, lower middle periapical radiographs; ULR, upper left and right periapical radiographs; LLR, lower left and right periapical radiographs; UMD, upper middle periapical radiographs; FMS, full-mouth series.

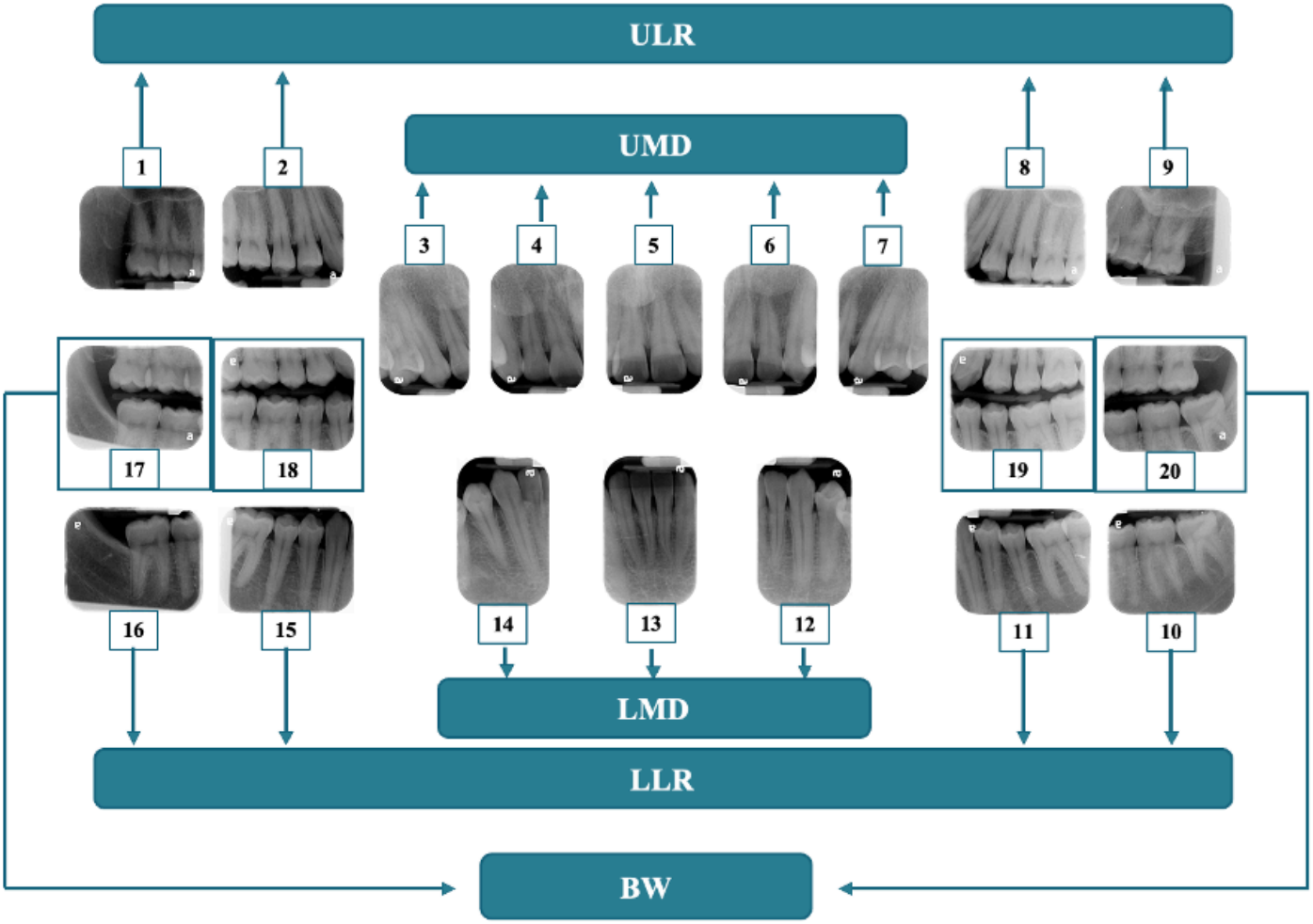


Figure 2

Spatial organization of FMS radiographs by Siamese network input group

Abbreviations: ULR, upper left and right periapical radiographs; UMD, upper middle periapical radiographs; LMD, lower middle periapical radiographs; LLR, lower left and right periapical radiographs; BW, bitewing radiographs.

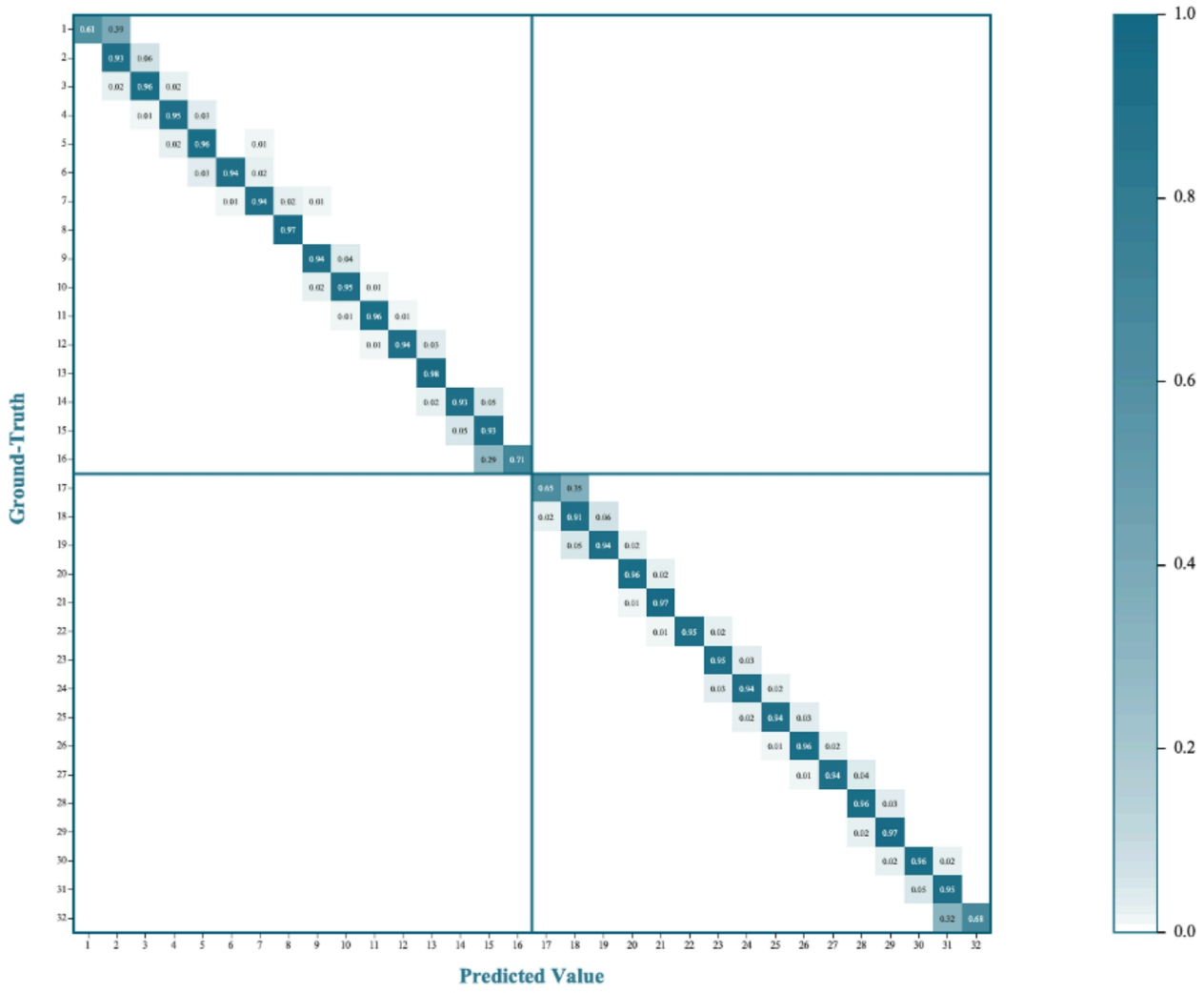


Figure 3

Confusion matrix of tooth-number prediction accuracy

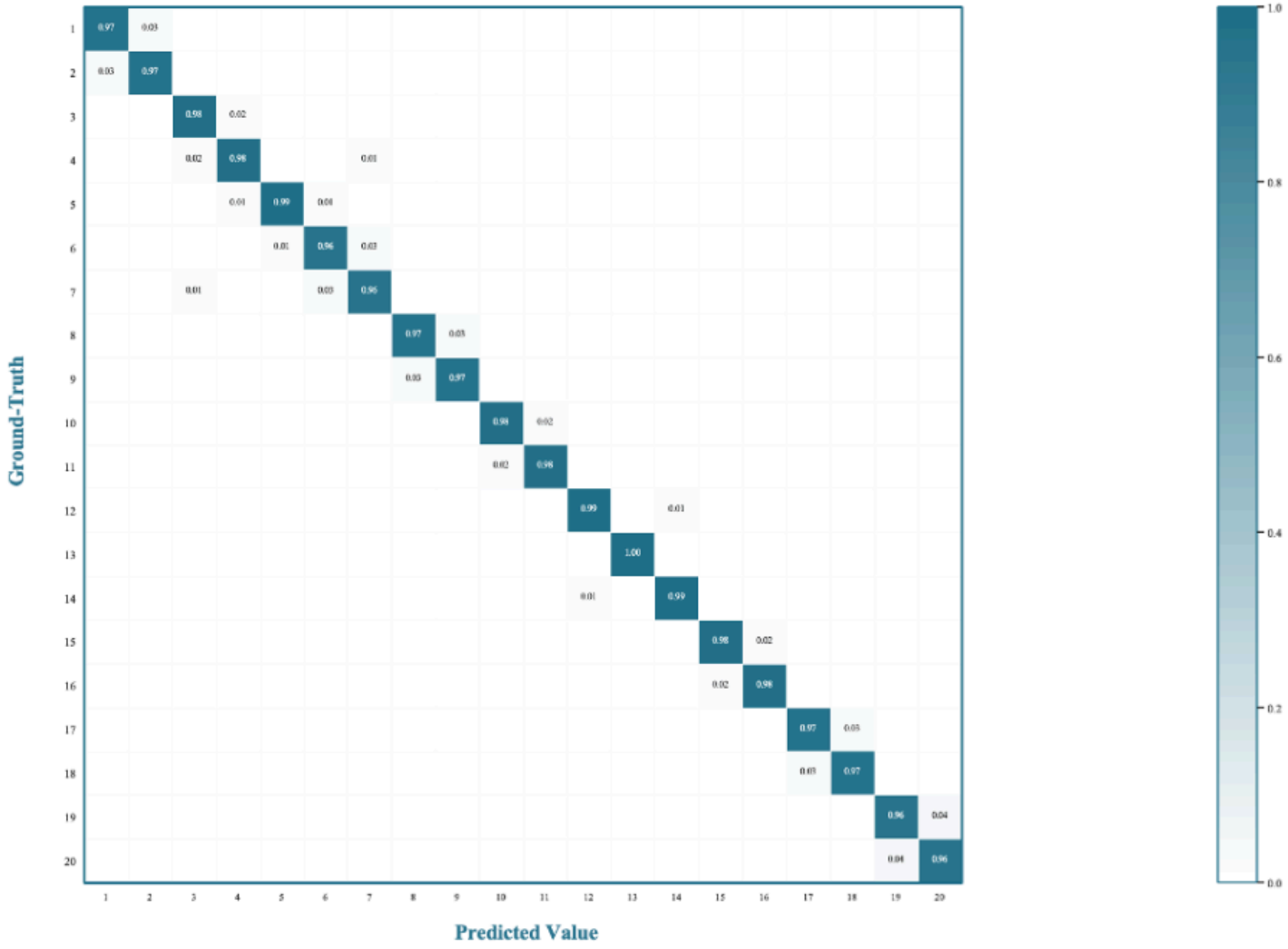


Figure 4

Confusion matrix of FMS placement accuracy