

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/189977/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Pachaiappan, Balaganesh, Poorni, Saravanan, Srinivasan, Srividhya, Nishanthine, Cruz, Srinivasan, Manali Ramakrishnan, Vinothkumar, Thilla Sekar, Nagendrababu, Venkateshbabu and Dummer, Paul M. H. 2026. Automated root canal curvature assessment from periapical radiographs using geometric and deep learning features. International Endodontic Journal 10.1111/iej.70281

Publishers page: <https://doi.org/10.1111/iej.70281>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



AUTOMATED ROOT CANAL CURVATURE ASSESSMENT FROM PERIAPICAL RADIOGRAPHS USING GEOMETRIC AND DEEP LEARNING FEATURES

Balaganesh Pachaiappan¹, Saravanan Poorni^{1}, Srividhya Srinivasan¹, Cruz Nishanthine¹, Manali Ramakrishnan Srinivasan¹, Thilla Sekar Vinothkumar², Venkateshbabu Nagendrababu^{3,4}, Paul MH Dummer⁵*

¹Department of Conservative Dentistry and Endodontics, Sri Venkateswara Dental College and Hospital, Chennai, India;

²Department of Restorative Dental Sciences, College of Dentistry, Jazan University, Jazan, Saudi Arabia;

³Department of Restorative Dentistry, College of Dental Medicine, University of Sharjah, Sharjah, UAE;

⁴Department of Conservative Dentistry and Endodontics, Dr. D. Y. Patil Dental College & Hospital, Dr. D. Y. Patil Vidyapeeth (Deemed to be University), Pimpri, Pune, India;

⁵School of Dentistry, College of Biomedical and Life Sciences, Cardiff University, Cardiff, UK.

Running title: Root Canal Curvature Assessment using Geometric and Deep Learning Features

Corresponding author: Saravanan Poorni, Department of Conservative Dentistry and Endodontics, Sri Venkateswara Dental College and Hospital, Chennai, India.

Email address: poorniis@yahoo.com

Funding: None received

Ethical statement: Ethical approval was obtained from the Institutional Review Board (IEC/SVDCH/2534)

Conflict of Interest: The authors have stated explicitly that there are no conflicts of interest in connection with this article.

Author contribution: All the authors made substantial contributions to the manuscript. All the authors have read and approved the final version of the manuscript.

Data availability statement: The datasets used and/or analyzed during the current study are

available from the corresponding author upon reasonable request.

Acknowledgment: The authors sincerely acknowledge Dr. Rakshna Nedumaran for her support and assistance in the preparation and illustration of Figure 1.

AI Declaration: The authors declare that Google Gemini Pro was used solely for language editing and improvement of the manuscript. The authors take full responsibility for the content of the work.

ORCID ID

Balaganesh Pachaiappan - 0009-0000-8831-3981

Saravanan Poorni - 0000-0001-8811-2960

Srividhya Srinivasan – 0000-0001-6929-4594

Cruz Nishanthine – 0000-0002-9709-8395

Manali Ramakrishnan Srinivasan - 0000-0002-3651-181X

Thilla Sekar Vinothkumar - 0000-0001-9798-3872

Venkateshbabu Nagendrababu - 0000-0003-3783-3156

Paul MH Dummer- 0000-0002-0726-7467

AUTOMATED ROOT CANAL CURVATURE ASSESSMENT FROM PERIAPICAL RADIOGRAPHS USING GEOMETRIC AND DEEP LEARNING FEATURES

ABSTRACT

Aim: To develop and validate a machine learning based automated computational framework using periapical radiographs for the detection, segmentation, and classification of root canal curvature in mandibular premolars with single canals, and to evaluate the contribution of geometric and deep learning features toward diagnostic performance.

Methodology: A retrospective cross-sectional study was performed using 1,644 periapical images of mandibular premolars. A four-stage multi-model framework was proposed using a single fixed 10-fold cross-validation partition applied consistently across all four stages to prevent data leakage between training and evaluation. The object detection models were YOLOv8, YOLOv11, and RF-DETR. For segmentation of the root canal, YOLOv8-seg, YOLOv11-seg, and Attention U-Net were used. The multiparametric feature extraction included five geometric morphometric features automatically derived from a binary mask, and 2,048 deep convolutional neural network features from ResNet-50. Classification of curvature was defined using an eXtreme Gradient Boosting (XGBoost) classifier with stratified 10-fold cross-validation with Synthetic Minority Over-sampling Technique (SMOTE) and algorithmic class weighting was applied within the training folds of each cross-validation iteration.

Results: The dataset included 633 straight, 885 moderately curved, and 126 severely curved canals. RF-DETR achieved a recall of 97.5%, precision of 93.3%, and F1-score of 95.4% for premolar detection. The YOLOv11-seg model was successful in segmenting the root canal with a Dice score of 85.8% and an IoU of 77.5%. All five geometric parameters were statistically significant for different severity groups (p -value <0.001). For the end-to-end curvature classification, the hybrid feature fusion model produced the highest overall metrics, achieving an accuracy of 89.60%, a recall of 83.80%, and an F1-score of 86.77%, compared to the models using either geometric features (88.38% accuracy) or deep learning features alone (82.42% accuracy).

Conclusions: The hybrid feature-fusion model reached the highest overall classification accuracy at 89.60%. Although, the geometric-only model had higher sensitivity for severe curvature with 80.6% compared to 69.1%, Lütten's angle was the most prominent predictor of curvature severity. Due to the limitations of two-dimensional radiographs in displaying buccolingual curvatures, this framework works best as a screening aid as it helps identify canals that need adjunctive three-dimensional evaluation.

KEYWORDS: Canal Curvature, Deep learning, Periapical radiography, Root canal, Endodontics

INTRODUCTION

The outcome of root canal treatment is determined by effective elimination of bacteria from the root canal system using a range of debridement and disinfection procedures. However, the process of debridement is complicated by the anatomical intricacies of root canal systems, including canal curvatures (Arias & Peters, 2022), which increase the risk of procedural errors including asymmetric dentine removal that results in canal transportation, ledges, zips and elbows, root perforations as well as instrument fractures (Chaniotis & Ordinola-Zapata, 2022). Though these mishaps are not always directly associated with the outcome of treatment, they tend to reduce the quality of canal debridement and the potential to remove pulp tissue, bacteria, and endotoxin (Lin et al., 2005). Consequently, canal curvatures are acknowledged as a risk factor that negatively influences the outcome of root canal treatment (Essam et al., 2021). Thus, preoperative evaluation of root canal anatomy is crucial, particularly the identification of factors such as the position of the beginning of the curvature along the root, the degree of curvature, and the radius of curvature (Sonntag et al., 2006). In line with established guidelines on endodontic case difficulty (American Association of Endodontists, 2018), the misclassification of a severely curved canal as a straight canal represents a critical diagnostic error, which immediately causes iatrogenic complications during biomechanical preparation. Iatrogenic canal mishaps are commonly associated with severe anatomical deformities when not detected preoperatively by practitioners (Hülsmann et al., 2005). This represents a critical error in accurately estimating canal curvatures, which may lead to inadequate biomechanical preparation of the apical third of canals, compromising chemical-mechanical debridement, resulting in a poor root canal filling characterized by the presence of voids, insufficient taper, and apical extension. Several studies have demonstrated a strong association between technically inadequate root fillings and significantly more post-treatment apical periodontitis, along with a reduced long-term survival rate of the treated teeth (Ng et al., 2011; Kirkevang et al., 2000). Severe curvatures may also increase the fatigue associated with nickel-titanium rotary instruments, which may cause instrument fracture when inappropriate file kinematics or non-flexible metallurgical file systems are used. Accurate

preoperative classification of canal curvatures of varying degrees is therefore critical in choosing appropriate file systems to avert such iatrogenic complications

Periapical radiography and cone beam computed tomography (CBCT) are routinely used to assess canal morphology pre-operatively. Using these imaging systems, several methods have been described to define root canal curvature, including Schneider's method (Zhu et al., 2003), Luiten's method (Luiten et al., 1995), Weine's method (Weine et al., 1982), and Cunningham's and Senia's method (Sonntag et al., 2005). Most of these established methods use periapical radiographs as the primary imaging modality for the purpose of diagnosis and treatment planning (Hilmi et al., 2023). However, in complex clinical situations, cone-beam computed tomography (CBCT) imaging increases diagnostic accuracy and facilitates more appropriate treatment strategies (Pinto et al., 2023). However, according to established clinical position statements on radiation safety (Special Committee to Revise the Joint AAE/AAOMR Position Statement on use of CBCT in Endodontics, 2015), CBCT is not routinely indicated for initial endodontic screening due to ionizing radiation constraints and increased economic cost (Sousa Melo et al., 2026; Patel et al., 2019).

Integrating automated image analysis of periapical radiographs into the clinical workflow offers advantages in clinical decision-making, with machine learning (ML) based algorithms providing a promising and scalable solution for the detection and analysis of root canal curvatures. These algorithms can be trained on large, annotated datasets to learn curvature patterns. Using Convolutional Neural Network (CNN), ML models can automatically segment the root canal space and calculate the angle of curvature (Kermany et al., 2018; Lahoud et al., 2021). Moreover, they provide real-time support for decision making by automatically identifying complex curvatures that require caution, helping clinicians select optimal file systems and instrumentation sequences tailored to each case. Integrated into digital radiographic software, machine learning-based tools offer a low-cost, non-invasive and time-efficient method that will enhance prognostic precision while reducing procedural errors, thus improving the treatment outcome (Khalifa & Albadawy, 2024).

Current clinical methodologies for assessing root canal curvature have clear deficiencies since they lack verifiable quantitative geometric parameters. Conversely, purely

mathematical algorithms offer these metrics but frequently fail due to radiographic image artefacts. Unfortunately, a more reliable methodology integrating both approaches is not available. Although several methods have been proposed for curvature assessment, most rely on manual evaluation (Rashid et al., 2015) or are applied to three-dimensional imaging modalities such as CBCT scans (Qiao et al., 2021; Thanaruengrong et al., 2021). In contrast, automated approaches tailored specifically for two-dimensional periapical radiographs are limited.

The current study aimed to develop and validate a machine learning-based automated computational framework utilizing archival patient records and periapical radiographs, for the detection, segmentation, and classification of root canal curvature in mandibular premolars with a single canal, and to evaluate the contribution of geometric and deep learning features toward diagnostic performance. The null hypothesis is: there is no statistically significant difference in classification performance when using geometric features alone, deep learning-derived features alone, or a hybrid combination of both for the assessment of root canal curvature.

MATERIALS AND METHODS

Study design

This study is a retrospective, cross-sectional, machine learning-based investigation utilizing archival patient records and periapical radiographs, aimed at developing and validating an automated computational framework for the assessment and classification of root canal curvature severity in mandibular premolars with a single canal.

Ethical approval, reporting guidelines

Prior to the commencement of the study, ethical approval was obtained from the local Institutional Review Board (IEC/SVDCH/2534) to ensure all data regarding human radiographs adhered to established ethical standards and regulatory guidelines. All data used in the study were anonymized, and no personal information was accessible at any stage of the project. The study is reported in accordance with the Preferred Reporting Items for Diagnostic Accuracy Studies in Endodontics (PRIDASE) as the primary reporting guideline

(Nagendrababu et al., 2024). Additionally, the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) was used as supplementary AI-specific reporting guideline (Tejani et al., 2024).

Data collection

A total of 1644 periapical radiographic images of mandibular premolars were retrieved retrospectively from the radiology database of the Department of Conservative Dentistry and Endodontics of Sri Venkateswara Dental College and Hospital, Tamil Nadu, India. The sample included all cases that met the inclusion criteria between July 2024 and July 2025. To ensure image consistency and eliminate inter-sensor variability, all images acquired from a single digital sensor were selected (Carestream RVG 5200; Carestream Dental, Atlanta, GA, USA), featuring a native resolution of 1168×1562 pixels and a $19 \mu\text{m}$ pixel pitch. The periapical radiographs were obtained for a variety of reasons not pertaining to the current research, including orthodontic, endodontic and surgical treatment, which were stored in the institutional radiology database.

The inclusion criteria for the radiographs were:

- Patients older than 18 years of age.
- Permanent mandibular premolar teeth that had not undergone endodontic treatment and exhibited complete root development with adequately visible canal morphology. A canal was considered adequately visible when its radiolucent path could be traced continuously from the canal orifice to the radiographic apex without interruption.
- Satisfactory quality of images in terms of contrast, sharpness and noise levels.
- Images providing sufficient clarity on root and canal curvatures.

The exclusion criteria were:

- Radiographs exhibiting severe distortion, motion artefacts, overlapping anatomical structures that obscured canal visualization, or incomplete visualization of the root when accurate expert annotation could not be achieved.

- Radiographs with artefacts obscuring the root canal space due to high-density restorative materials (e.g., metallic crowns, amalgam, or other radiopaque restorations).

Data pre-processing

In order to maintain consistency across samples, all the images procured were resized to a standard resolution of 640 x 640 pixels using bilinear interpolation. This specific resolution was selected as it is the optimal architectural input requirement for the object detection and segmentation models, balancing computational efficiency with the retention of spatial features. The images were normalized by scaling intensity values to the [0,1] range. To improve the visibility of the root canals, contrast enhancement was performed using Contrast Limited Adaptive Histogram Equalization (CLAHE) applied using standard parameters (clip limit = 2.0, tile grid size = 8x8). To account for spatial deformation introduced by resizing the rectangular radiographs (original resolution 1168 × 1562 pixels; native pixel pitch of 0.019 mm) to the 640 × 640 pixel model input, all automated geometric outputs were mathematically resized relative to the original sensor dimensions prior to final calculation. In response to non-uniform dimensional compression during the pre-processing stage, horizontal scaling factors ($S_x = 1.825$) and vertical scaling factors ($S_y = 2.441$) were integrated to effectively map the standardized 640 × 640 pixel coordinates into the native sensor matrix, ensuring all linear measurements were invariant. The scaling constants were calculated as: $S_x = \text{Original Width} / \text{Scaled Width}$; $S_y = \text{Original Height} / \text{Scaled Height}$.

To prevent data leakage, a unified, stratified 10-fold cross-validation framework was implemented across the entire experimental pipeline. The complete dataset ($n = 1,644$) was iteratively partitioned, with each fold allocating 90% of the images for model training and sequestering 10% exclusively for testing to ensure no interaction occurred between the sets. Within each training fold, a nested 10% validation split was used for hyperparameter tuning and early stopping. Finally, to increase dataset variability and prevent overfitting, dynamic data augmentation including random rotations ($\pm 15^\circ$), horizontal flipping ($p = 0.5$), Gaussian noise addition ($\sigma = 0.01$), and brightness adjustments ($\pm 20\%$) was applied solely to the training folds during each cross-validation iteration.

A single, fixed stratified 10-fold partition of the complete dataset was generated at the image level, preserving the proportional representation of straight, moderate, and severe curvature classes in each fold. This same fold assignment was held fixed and applied identically across all four pipeline stages: premolar detection, root canal segmentation, feature extraction, and curvature classification. For each fold k , models at every stage were trained exclusively on the images from the remaining nine folds and evaluated only on fold k . Critically, evaluation of the held-out fold at each stage used the predicted output of the preceding stage, not its ground-truth annotation: the bounding box used to crop the premolar region was the detection model's own prediction; the binary mask used for feature extraction was the segmentation model's own prediction on that cropped region; and the resulting geometric and CNN features were passed to the classifier trained on the other nine folds. Reported performance therefore reflects the accuracy of the complete, fully automated four-stage pipeline, including any error propagated from upstream stages, rather than an idealized scenario using ground-truth intermediate outputs.

Experimental pipeline

The semi-automated image analysis framework was constructed as a progressive four-stage multi-model computational pipeline: (1) Premolar detection, (2) Root canal segmentation, (3) Multiparametric feature extraction (geometric and deep learning), and (4) curvature classification. All the computational tasks are conducted on the Kaggle platform (Google LLC, Mountain View, CA, USA), a cloud-based workstation equipped with two NVIDIA T4 GPUs (Nvidia Corporation, Santa Clara, CA, USA) with 30GB VRAM.

Annotation and Ground Truth Establishment

Prior to model development, manual annotations were performed for all 1,644 periapical radiographs using Roboflow (Roboflow Inc., Des Moines, IA, USA) to generate the reference standard for object detection, segmentation, and curvature classification. Two experienced endodontists, with 13 and 15 years of clinical experience in endodontics and radiographic interpretation, performed all annotation procedures independently.

For object detection, bounding boxes were manually drawn around the premolar region of interest (ROI). The cropped ROI was subsequently used for segmentation, where the operators manually delineated the root canal pathway using polygon annotations that accurately represented the canal morphology. These manually annotated polygon masks served as the ground truth for training and evaluating the segmentation models.

To establish the reference standard for curvature classification, the manually annotated canal masks were digitally superimposed onto their corresponding original periapical radiographs using a semi-transparent overlay implemented in OpenCV (Open Source Computer Vision Library; <https://opencv.org>) and Matplotlib (Matplotlib Development Team; <https://matplotlib.org>), preserving visualization of both the segmented canal and the surrounding anatomical structures. The two endodontists independently classified each canal as straight, moderately curved, or severely curved according to Schneider's classification. Curvature assessment was performed exclusively on the internal root canal trajectory, rather than the external root contour, consistent with Schneider's original description of canal curvature. While conventional thresholds define straight canals as $\leq 5^\circ$ and moderate canals as $10\text{--}20^\circ$, anatomical variations falling within the $6^\circ\text{--}9^\circ$ transitional range were resolved through expert clinical judgment. Canals within this gap exhibiting negligible visual deviation that would not necessitate altered instrumentation protocols were classified as "straight".

Disagreements between the two observers were resolved through consensus meetings. Cases in which consensus could not be reached were reviewed by a third senior endodontist, whose decision was considered final. Inter-observer agreement for curvature classification was evaluated using Cohen's kappa statistic (IBM SPSS Statistics, Version 31.0, Armonk, New York, USA), demonstrating almost perfect agreement ($\kappa = 0.82$). The resulting consensus labels constituted the final ground truth for all subsequent classification experiments.

To ensure an unbiased evaluation, these expert-derived reference labels were established before model development and remained fixed throughout the study. During model evaluation, geometric and deep-learning features were extracted from the predicted segmentation masks generated within each cross-validation fold, whereas the reference

class labels remained those assigned from the expert-annotated masks. This strategy ensured complete separation between ground-truth label generation and model prediction during performance evaluation.

Stage1: Premolar detection

To localize the mandibular premolar Region of Interest (ROI), three established object detection models were assessed: YOLOv8, YOLOv11 (Ultralytics Inc., Los Angeles, CA, USA), and RF-DETR (Roboflow Inc., Des Moines, IA, USA). YOLOv8 and YOLOv11 were initialized with COCO-pretrained weights and trained using their standard loss formulas (Complete Intersection over Union [CIoU] and Distribution Focal Loss [DFL] for bounding-box regression; Binary Cross-Entropy [BCE] for classification), applying cosine learning-rate decay. In contrast, RF-DETR was initialized from medium-variant pretrained weights. It used a DINOv2-windowed-small vision transformer encoder (12 layers, hidden dimension 256, patch size 16). Notably, since this patch size differs from DINOv2's default size of 14, only the RF-DETR checkpoint weights were loaded, separate from the original DINOv2 backbone. This setup was paired with a 4-layer group-DETR decoder (13 groups, 300 queries, two-stage detection with bounding-box reparameterization; 33.4M total parameters). Following a unified 10-fold cross-validation protocol, all models were optimized with an AdamW (Adaptive Moment Estimation with Weight Decay) optimizer. RF-DETR was trained with a fixed learning rate (0.0001 for the backbone, 0.00015 for the encoder), using an exponential moving average of model weights (decay = 0.993) for final evaluation. To avoid overfitting and ensure good generalization, early stopping mechanisms were applied across all training folds. RF-DETR was ultimately chosen as the main detection model for the primary pipeline due to its higher recall. In this sequential system, missing a premolar detection results in a total downstream failure for that specific case. On the other hand, a detection with less precise localization (affecting mAP@0.5:0.95) could still be successfully captured. The best-performing model automatically generated the cropped ROI, which served as the localized input for the downstream segmentation phase. To keep size consistency for these varying bounding boxes, adaptive zero-padding was applied, embedding the cropped ROI into a fixed 640×640 pixel canvas with zero-value peripheral pixels. This padding strategy preserved

the native sensor pixel pitch and avoided uneven geometric stretching, thus maintaining the true anatomical paths needed for accurate downstream mathematical angle calculations.

Stage 2: Root canal segmentation

Following the successful localization of the premolar ROI, precise segmentation of the root canal space was performed to create a binary mask that isolated the canal from the surrounding tooth structure. Three segmentation architectures were compared: YOLOv8-seg, YOLOv11-seg, and Attention U-Net (Al Qurri & Almekkawy, 2023). All three models were trained using an AdamW optimizer, which was the same as in the detection stage. The YOLO segmentation models started with COCO-pretrained weights (Ultralytics defaults) because of their high efficiency in instance segmentation. They were trained for up to 200 epochs, with a batch size of 16 and a learning rate of 0.01, using their default instance-segmentation loss, which includes combined box, classification, and mask losses. Attention U-Net was trained for up to 100 epochs, with a batch size of 4 and a base learning rate of 0.0001, using a combined cross-entropy and Dice loss. Early stopping was applied to all models to prevent overfitting. Segmentation performance was evaluated using the 10-fold cross-validation method mentioned earlier. Metrics included the Dice similarity coefficient (DSC), IoU, cIDice, 95th-percentile Hausdorff Distance (HD95), Average Symmetric Surface Distance (ASSD), precision, and recall.

Stage 3: Multiparametric Feature extraction (Geometric and Deep learning)

To comprehensively quantify the morphological severity of the root canal curvature, a dual feature extraction pipeline was constructed. This framework was designed to extract features from the binary masked root canal outputs generated by the best-performing segmentation algorithm and transform them into a multiparametric profile. First, each binary mask was improved with a 3×3 morphological closing operation to remove small gaps. Then, connected-component analysis was performed to keep only the largest connected area that matched the root canal. Next, a continuous centreline of the canal was extracted using a scanline-based midpoint algorithm. This algorithm calculated the midpoint between the leftmost and rightmost mask pixels along the main canal axis while leaving out

disconnected or branching rows. The extracted centreline was automatically aligned from the canal opening to the apex using a Euclidean distance transform, which presumed a wider canal at the coronal end. To reduce artefacts near the pulp chamber, a dynamic trimming method removed unnecessary coronal projections. After this, the centreline was smoothed with a moving-average filter (window size = 25 points) and fitted with a cubic B-spline (degree = 3). The cumulative arc length of the spline was then calculated and converted to the native image resolution through bidirectional coordinate scaling. The processed centreline became the common geometric representation for calculating Lütten's Angle, Pruett's Radius, Canal Access Angle, Curvature Height, and Curvature Distance.

To account for spatial deformation introduced by resizing the rectangular radiographs for model input, all automated geometric outputs were mathematically rescaled relative to the original sensor dimensions (1168×1562 pixels; native pixel pitch of 0.019 mm) prior to final calculation. Utilizing this calibrated spatial conversion, the following five continuous morphometric parameters were extracted for each canal as shown in Figures 1 & 2:

Lütten's Angle (Modified Schneider method): The acute angle formed by the intersection of the defined coronal and apical central axes (Luiten et al., 1995).

Pruett's Radius: the radius of the circle that tangentially adheres to the canal at the point of maximum curvature (Pruett et al., 1997).

Canal Access Angle (CAA): the angle between the long axis of the tooth and the line connecting the access cavity entry point to the apex (Günday et al., 2005).

Curvature Height: The maximum perpendicular distance (in millimetres) from a line drawn between the canal orifice and the apex to the point of maximum curvature on the canal curve (Günday et al., 2005).

Curvature Distance: The longitudinal distance (in millimetres) traced along the orifice-apex line to the point of maximum curvature height (Günday et al., 2005).

These five parameters provided a mathematical representation of the canal geometry, replicating manual endodontic assessment while remaining robust against segmentation noise.

To capture complex spatial patterns that could not be adequately represented by geometric assessment alone, a pretrained CNN model (ResNet-50) was utilized for deep feature extraction. Since the model requires three-channel input (RGB), the one-channel binary mask was duplicated into three channels to simulate RGB images. To utilize the model only for feature extraction rather than object classification, the final fully connected layer was replaced with an identity mapping layer, this forced the model to produce raw spatial embeddings directly from the global average pooling layer, yielding 2048-dimensional features for each binary image. Following the independent extraction protocol, the multiparametric dataset was created for each segmented binary image. Both geometric parameters ($n = 5$) and deep learning parameters ($n = 2048$) were concatenated based on unique image identifiers. This feature fusion generates a multiparametric profile that contains 2053 unified features per canal mask.

Stage 4: Curvature classification

To evaluate the diagnostic contribution of both geometric and deep learning features, an ablation study was designed. eXtreme Gradient Boosting (XGBoost; Distributed Deep Machine Learning Community; <https://xgboost.ai>) was selected as the primary classifier due to its efficiency in structured feature classification tasks. The model was trained and evaluated in three different setups: with only 5 geometric features, only the 2048 CNN features, and all 2053 hybrid features. To prevent data leakage, a stratified 10-Fold Cross-Validation was used to ensure the model was tested entirely on unseen data. Furthermore, to address the inherent class imbalance within the dataset (particularly the minority 'Severe' class) and prevent the algorithm from developing majority-class bias, the Synthetic Minority Over-sampling Technique (SMOTE) combined with algorithmic class weighting was applied within the training folds of each cross-validation iteration. Performance was measured using Accuracy, Precision, Recall, and F1-Score. Additionally, the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) and the Precision-Recall Curve (PR-AUC) were

calculated to strictly validate the model's discriminative thresholding and resilience to the minority class. Classification performance was evaluated at two levels. First, conditional classification performance was assessedⁱ for radiographs that successfully passed the upstream premolar-detection stage. Second, end-to-end pipeline performance was assessed using all radiographs, with upstream premolar-detection failures recorded as missed-detection (MD) outcomes. These cases were not assigned a downstream curvature class and were incorporated into the end-to-end performance assessment. The end-to-end results, which included both successful classifications and upstream detection failures, were used as the final performance results of the proposed pipeline. This ablation framework allowed systematic evaluation of the relative contribution of geometric and CNN-derived features to curvature severity classification.

To interpret XGBoost decision-making and evaluate the clinical relevance of extracted parameters, a formal feature importance analysis was conducted. Feature importance ranking was calculated via information gain were extracted directly from the classifier across both isolated and hybrid models. This allowed transparent evaluation of which parameters influenced final root canal curvature severity, ensuring the model diagnostic logic aligned with established endodontic principles. Supplementary Figure S1 shows the schematic overview of the experimental multi-model computational pipeline. The automated framework consists of (1) Input, (2) premolar detection, (3) root canal segmentation, (4) multiparametric feature extraction combining geometric and deep learning parameters, and (5) final root canal curvature classification. The data flow diagram is represented in Supplementary Figure S2.

Statistical Analysis

Data was analysed using SPSS Statistics software (Version 31.0; IBM Corp., Armonk, NY, USA) and Python (Python Software Foundation, Wilmington, DE, USA). Inter-examiner reliability for the ground truth categorical annotations was measured using Cohen's kappa coefficient. The continuous geometric variables (Luiten's angle, Pruett's radius, canal access angle, curvature height, and curvature distance) were initially assessed for normal distribution. One-way analysis of variance (ANOVA) was performed to compare these geometric

parameters. For pairwise comparisons, the Tukey honestly significant difference (HSD) test was applied. Performance for object detection and segmentation was quantified using mAP, DSC, IoU, precision, recall, and F1-score. The predictive performance of the final XGBoost classification framework was validated using stratified 10-fold cross-validation. The significance level was adjusted at $p < 0.05$.

RESULTS

Dataset distribution

A total of 1,644 retrospective periapical radiographic images were visually analysed by two experienced endodontists. The overall dataset included 633 straight canals, 885 moderately curved canals, and 126 severely curved canals as summarised in Table 1. A unified, stratified 10-fold cross-validation framework was utilized across all computational stages, encompassing object detection, segmentation, and classification. Ground-truth annotations used for classification were independently annotated by two experienced endodontists holding a master's degree with fifteen and thirteen years of clinical and academic experience, respectively. Inter-rater reliability analysis demonstrated almost perfect agreement (Cohen's $\kappa = 0.82$). Any discrepancies between the annotations were resolved through consensus.

Object detection performance

Three object detectors were evaluated for premolar localization (Table 2). RF-DETR achieved the highest recall ($97.5 \pm 0.8\%$) and F1-score ($95.4 \pm 0.4\%$), indicating higher sensitivity for detecting premolars while maintaining overall detection performance. YOLOv8 demonstrated the highest precision ($93.5 \pm 1.2\%$), mAP@0.5 ($98.3 \pm 1.4\%$), and mAP@0.5–0.95 ($91.4 \pm 1.3\%$), reflecting the most accurate localization performance across both standard and stricter IoU thresholds. YOLOv11 showed balanced performance, with a precision of $91.6 \pm 1.1\%$, recall of $93.8 \pm 1.3\%$, and F1-score of $92.7 \pm 1.0\%$, although its localization performance at higher IoU thresholds (mAP@0.5–0.95 = $86.2 \pm 1.0\%$) was lower than that of YOLOv8.

Segmentation performance

Following extraction of the premolar ROI, three deep learning models were evaluated for root canal segmentation. YOLOv11-seg achieved a Dice score of 85.87%, IoU of 77.53%, cIDice of 90.36%, HD95 of 0.5813 mm, and ASSD of 0.159 mm. It also achieved the highest precision (89.24%) and recall (87.57%). Attention U-Net recorded a marginally higher Dice score (86.17%) than YOLOv11-seg, but showed lower cIDice, and higher HD95 and ASSD. YOLOv8-seg was outperformed by YOLOv11-seg across all seven metrics, including ASSD (0.1992 mm vs. 0.159 mm). As downstream geometric measurement (Stage 3) depends on skeleton and boundary precision rather than raw pixel overlap alone, YOLOv11-seg was selected as the operational segmentation model for the pipeline (Table 3).

Geometric feature analysis

Samples of automated geometric assessment of a straight, moderate and severe root canals are presented in Figures 3, 4, and 5 respectively. One-way ANOVA revealed significant differences ($p < 0.001$) across all five geometric parameters when comparing straight, moderate, and severe canal groups (Table 4). Pairwise post-hoc comparisons are presented in Supplementary Table S1. As visually demonstrated by the morphological distributions (Figure 6), Luiten's angle, Canal Access Angle, and Curvature Height were associated with the largest effect sizes ($\eta^2 = 0.26$ – 0.59) and increased proportionally with severity; Pruett's Radius ($\eta^2 = 0.203$) decreased with severity. Curvature Distance showed a comparatively small effect ($\eta^2 = 0.010$) and, subsequent post-hoc pairwise comparisons confirmed that while Curvature Distance successfully differentiated Straight from Moderate canals ($p < 0.001$), it failed to significantly distinguish Severe canals from either the Straight ($p = 0.183$) or Moderate ($p = 0.927$) cohorts, indicating that this specific parameter contributes little discriminative diagnostic value for identifying severe curves.

Classification performance

Table 5 presents the 10-fold cross-validation end-to-end classification results for the XGBoost model under three feature conditions. The model utilizing exclusively geometric features ($n = 5$) returned an accuracy of 88.38% ($\pm 1.77\%$), a macro-precision of 87.28%, a

macro-recall of 86.30%, and a macro-F1-score of 86.61%. The model restricted to CNN features (n = 2048) produced a lower accuracy of 82.42% ($\pm 2.19\%$), with a macro-precision of 82.50%, a macro-recall of 70.49%, and a macro-F1-score of 73.87%. The hybrid feature fusion dataset (n = 2053) yielded the highest overall metrics, recording an accuracy of 89.60% ($\pm 2.11\%$), a macro-precision of 91.63%, a macro-recall of 83.80%, and a macro-F1-score of 86.77%. These end-to-end results include the 41 radiographs with upstream premolar-detection failures, which were retained as missed-detection (MD) outcomes. The corresponding conditional classification results among the 1,603 successfully detected radiographs are provided in Supplementary Table S2. Per-class precision, recall, and performance of the classification models across each curvature is summarized in Table 6; the corresponding ROC and precision–recall curves for each class (Figure 7) and To interpret the algorithmic decision-making process and validate its clinical relevance, XGBoost feature importance rankings were extracted across the ablation framework (Figure 8). In the isolated geometric model, Lüten’s angle emerged as the absolute dominant predictive parameter, contributing 52.6% to the model's overall decision weight. This was sequentially followed by Pruett’s radius (18.1%), Curvature Height (13.1%), Canal Access Angle (9.4%), and Curvature Distance (6.8%). Notably, even when evaluated against 2,048 deep learning features within the hybrid fusion model, explicit geometric parameters specifically Lüten’s angle and Curvature Height remained the primary algorithmic discriminators. This algorithmic prioritization confirms that the framework relies heavily on verifiable clinical geometry, rather than abstract spatial noise, to accurately categorize root canal curvature severity.

Confusion Matrix: The diagnostic performance of the geometric-only, CNN-only, and hybrid XGBoost classification models was evaluated with 10-fold cross-validation and is summarized in the confusion matrices in Figure 9. Analysis of these misclassifications reveals a distinct directional pattern across all three patterns: the algorithms demonstrated a strong capacity to differentiate between morphological extremes, with confusion between straight and severe canals remaining minimal throughout. The geometric-only model misclassified 5 straight canals as severe and 3 severe canals as straight; the CNN-only model misclassified 3 straight canals as severe and 7 severe canals as straight; and the hybrid model

misclassified 2 straight canals as severe and 2 severe canals as straight. The vast majority of all diagnostic discrepancies occurred exclusively between adjacent categories. Errors predominantly involved classifying borderline moderate curvatures as straight, or misidentifying severe curvatures as moderate. This distribution pattern suggests that the measurement discrepancies closely mirror the continuous, gradual nature of anatomical root curvature, rather than demonstrating random categorization. In addition to these classification errors, 41 radiographs were recorded as missed detections at the upstream premolar detection stage, comprising 14 straight, 21 moderate, and 6 severe cases. These missed detections were retained in the end-to-end analysis rather than being excluded.

DISCUSSION

The primary objective of this retrospective study was to develop and validate a standardized, automated diagnostic tool to evaluate the severity of root canal curvature on two-dimensional periapical radiographs. Mandibular premolars present significant morphological complexities including curvatures, apical bifurcations, trifurcations and C-shaped canal variations (Thanaruengrong et al., 2021). This anatomical variability makes accurate preoperative assessment both challenging and critical. Statistical analysis of the extracted geometric parameters demonstrated significant variations across the three clinical difficulty classes (straight, moderate, and severe). Moreover, the proposed integrated approach, which combined both geometric measurements and automated image analysis using deep learning, significantly outperformed standalone deep learning models. The success of this integrated mathematical assessment was fundamentally driven by the accurate automated segmentation phase (YOLOv11-seg), which reliably isolated the anatomical pathways with a Dice score of 85.8% and an IoU of 77.53%. Due to successful object detection and segmentation, the combined diagnostic model achieved an overall accuracy of $89.60 \pm 2.11\%$ and a macro-averaged F1-score of 86.77 %. Therefore, based on these strong clinical and statistical findings, the null hypotheses were rejected.

The YOLOv11-seg achieved a Dice Similarity Coefficient (DSC) of 85.8% for root canal segmentation. Contextualizing this metric is challenging, as existing two-dimensional (2D) literature primarily focuses on whole-tooth or lesion segmentation, while commercial

diagnostic platforms utilize closed, proprietary datasets and do not disclose pixel-level spatial overlap metrics (Kazimierczak et al., 2024; Schwendicke et al., 2020). Consequently, the most relevant academic benchmarks are three-dimensional (3D) cone-beam computed tomography (CBCT) studies. Recent state-of-the-art deep learning models operating on 3D CBCT volumes have achieved Dice scores exceeding 96% for whole-tooth segmentation (Chen et al., 2025); however, when evaluations are restricted specifically to the internal pulp cavity and root canal systems in isolation, instance-level Dice scores drop ranging from 88%-93% even when utilizing precise micro-CT derived ground truth (Slim et al., 2024), down to approximately 62-65% in broad multi-structure challenges (Wang et al., 2025). Against this standard, the present model maintained highly effective spatial localization despite the inherent challenges of 2D anatomical noise and superimposition. Boundary deviations, quantified by an HD95 of 0.581 mm and an ASSD of 0.159 mm, reflect minor spatial inconsistencies. Crucially, while post-hoc smoothed spline fitting successfully mitigated jagged trajectories, residual CNN boundary noise can still cause apical-landmark localization errors when the canal terminus is obscured by dense trabeculation. These minimal spatial shifts occasionally alter borderline vector slopes, contributing to the remaining classification challenges at the extreme margins of curvature severity.

This study represents a novel approach, that combines mathematical assessment with automated image analysis using established endodontic principles. An alternative to a learned classifier would be a deterministic geometric threshold rule applied directly to segmented structures, similar to approaches using topological analysis to derive clinical indices from segmented dental anatomy (Bui et al., 2023). We opted for a supervised classifier instead, because canal severity reflects the joint behaviour of several correlated geometric parameters rather than a single measurable distance, and a data-driven model can learn this joint boundary more robustly than a fixed univariate cut-off, particularly near the moderate-severe transition. While conventional deep learning models have demonstrated utility in dental imaging, their diagnostic performance is often limited when evaluating complex anatomical variations solely on two-dimensional periapical radiographs (Mohammad-Rahimi et al., 2024). Standard computational models rely predominantly on identifying recognizable pixel-based patterns from dataset distributions rather than

possessing genuine anatomical comprehension (Ourang et al., 2025). Consequently, when evaluating inherently imbalanced endodontic morphologies and extreme-case difficulties, these classifier networks struggle to generalize, optimizing for the majority anatomical classes and resulting in inadequate sensitivity to severe deviations (Karkehabadi et al., 2024). This limitation was directly validated in the baseline comparison, where the standalone image-based model yielded a macro-recall of only 70.49%. To overcome the inherent vulnerability of strictly pixel-based learning, the present methodology integrated explicit mathematical invariants, specifically Luetin's angle, curvature height and Pruett's radius. This multiparametric feature fusion forced the terminal classifier to mathematically acknowledge extreme morphological dilacerations, functionally replicating the clinical reasoning and prior anatomical knowledge utilized by human experts (Ourang et al., 2025). By embedding these established geometric principles directly into the diagnostic tool, the recall was successfully elevated to 83.80%. Most importantly, to address the extreme class imbalance inherent to current dataset, this framework integrated the SMOTE and algorithmic class weighting strictly within the training folds of a stratified 10-fold cross-validation protocol. This methodological optimization also directly solved the algorithm's prior vulnerability to majority-class bias, propelling the model to achieve higher sensitivity (recall) for severely curved root canals, resulting in a balanced classification system that circumvents the clinical limitations of standard automated image analysis.

To assess how each variable affects model performance, we calculated XGBoost feature importance rankings. In the geometric model, Lütten's angle was the main predictor, making up 52.6% of the decision weight. When we compared it to 2,048 deep learning features within the hybrid framework, explicit clinical metrics specifically Lütten's angle and Curvature Height remained the top discriminators. This indicates that the model values measurable clinical geometry more than random patterns in determining root canal curvature severity. This supports the clinical validity of the automated output.

A major challenge in automated canal segmentation is the nature of binary masks generated by convolutional models. Standard skeletonization often creates jagged edges because of pixel quantization. As a result, small deviations along the extracted path can cause instability

in vector slopes, leading to inconsistent geometric calculations. To overcome this, the current system used spline smoothing to turn the discrete pixel centreline into a continuous path. By basing the calculations of Lütten's Angle and Pruett's Radius on this optimized spline instead of the raw pixel data, the system made sure that the resulting metrics stayed consistent. This approach reduced the impact of minor boundary noise from the initial segmentation and improved the mathematical reliability of the resulting geometric parameters.

A primary limitation of this study is its reliance on two-dimensional periapical radiographs. Standard 2D projections restrict the evaluation of root canal curvature to the mesiodistal plane, creating a diagnostic limitation regarding buccolingual curvatures (Albitar et al., 2022). Because both the model input and the ground-truth references were constrained to 2D projections, the algorithm cannot quantify anatomical variations occurring perpendicular to the image sensor.

Consequently, this framework should be viewed as an adjunctive clinical triage tool rather than a replacement for three-dimensional imaging. When the algorithm identifies severe curvature or detects ambiguous anatomical features, it serves as an objective indicator for the clinician to consider a Cone Beam Computed Tomography (CBCT) scan. Integrating this triage protocol ensures that volumetric imaging is reserved for complex cases where it is clinically indicated, thereby optimizing diagnostic efficiency in accordance with the ALARA principle. This staged approach is consistent with current clinical guidance indicating that CBCT is not warranted for initial routine endodontic screening (Sousa Melo et al., 2026; Patel et al., 2019) but remains valuable as a secondary reference standard once a case has been flagged for closer evaluation.

The hybrid framework achieved 69.10% ($\pm 17.02\%$) sensitivity for severe dilacerations, reducing clinically important false negatives compared with the CNN-only model (40.58% $\pm 17.56\%$) but a small number of severe canals were misclassified as straight. Review of these cases revealed two major causes. First, the visibility of the canal path was reduced by segmentation inaccuracies due to radiopaque anatomical superimposition, such as dense trabecular bone and localized sclerosis, especially in the apical third. This caused incomplete

segmentation of the canal and underestimation of the canal curvature in the extraction of geometric features. Smoothing in post-processing improved irregular mask boundaries, but it was not able to recover canal segments that were not correctly segmented. Second, borderline cases with curvature measurements close to the predefined classification threshold ($>20^\circ$ for severe) were sensitive to small variations in the extracted geometric features and resulted in misclassification of neighbouring severity categories. These results indicate that the majority of classification errors were caused by upstream segmentation limitations and the use of fixed geometric thresholds, rather than the classification model. Thus, we expect that improving apical canal segmentation will enhance the performance of the proposed framework.

Whereas existing dental image analysis research has begun to increasingly incorporate CBCT modalities in the evaluation of morphological intricacy, the present framework has exclusively relied upon two-dimensional periapical radiographs. Recent investigations have successfully applied spatial architectures to three-dimensional volumetric data for anatomical segmentation (Duman et al., 2024), as volumetric imaging inherently eliminates the superimposition of adjacent anatomical structures and provides a high-fidelity signal that facilitates robust computational feature extraction (Patel et al., 2007).

First, the dataset was acquired using a single digital sensor (Carestream RVG 5200), by a single operator, at a single institution. While this standardization minimized external variance during validation of the geometric formulae, it inherently limits generalizability. Domain shift across imaging sources remains a well-documented challenge in dental radiographic AI, as different sensors produce distinct baseline noise profiles, contrast gradients, and spatial. Future work should address this through multi-sensor training datasets, domain adaptation, transfer learning, and knowledge distillation to achieve hardware-agnostic performance.

Second, the inclusion criterion of satisfactory radiographic image quality introduced a selection bias. By excluding lower-quality images and those with substantial artefacts, the dataset systematically omitted the most obscured, diagnostically challenging cases, precisely the scenarios in which an automated diagnostic aid would arguably offer the greatest clinical

benefit. Additionally, demographic data beyond the ">18 years" criterion were not recorded, precluding analysis of potential age-related calcification patterns or sex-based morphological differences.

Third, although the XGBoost feature-importance analysis in which Lütten's angle ranked first among 2,053 features supports the model's technical interpretability, clinical explainability was not formally assessed as an outcome measure. Demonstrating algorithmic logic is not equivalent to establishing clinical explainability for end-users; prospective reader studies are needed to evaluate how practitioners interpret these automated geometric outputs and whether they influence treatment planning.

Finally, the computational framework was limited to internal canal geometry. Although internal canal trajectory is the primary determinant of bending stress in endodontic instruments, integrating internal canal geometry with external root morphology could provide a more comprehensive assessment of root anatomy, representing a logical direction for future multi-parametric AI models.

Future research needs to validate the suggested framework through multi-centre datasets from various radiographic devices. The quality of images, the properties of the sensors, and the techniques of acquiring images may be key to the performance of the model. Therefore, domain adaptation, transfer learning, and knowledge distillation techniques can help improve consistency between the imaging platforms and aid in implementing the framework successfully. Moreover, an important area for further study is the need for reader studies to compare the performance of clinicians who receive assistance from the framework and those who do this independently. Additionally, a granular, parameter-by-parameter 1:1 comparison between observer-measured geometric parameters and the model's outputs is warranted to provide deeper diagnostic validation. On the anatomical side, future studies should consider not only single-rooted premolars but involve multi-rooted teeth, particularly molars characterized by complicated root morphology and canal systems. The investigative efforts should also employ the three-dimensional cone-beam computer tomography (CBCT) data in further analyses as the investigation had been carried out on the basis of two-dimensional radiographic images.

Despite these promising findings, the clinical application of the framework remains constrained by the inherent limitations of two-dimensional radiography, which cannot adequately depict buccolingual curvatures, and by its dependence on accurate automated segmentation, which may be affected by anatomical superimposition. Furthermore, the present study was limited to mandibular single-rooted premolars, and therefore its findings cannot be directly generalized to other tooth groups. Consequently, the proposed framework is not intended to replace clinical judgement, but rather to serve as an adjunctive decision-support tool that objectively identifies severe canal curvatures and assists clinicians in determining when additional three-dimensional cone-beam computed tomography (CBCT) imaging may be indicated. Further validation using multi-centre datasets and more complex multi-rooted teeth is required before routine clinical implementation.

CONCLUSION

The current study has demonstrated the feasibility of an automated hybrid computational framework for categorising the severity of root canal curvatures from two-dimensional periapical radiographs. The hybrid feature-fusion model attained the overall highest classification accuracy, exceeding geometric and deep learning features employed independently. Among the geometric parameters, Luiten's Angle was the most significant predictor of severity. These findings assist the potential of integrating clinically definable geometric features with deep learning for automated preoperative assessment of root canal curvature and as an additive decision-support tool in endodontic treatment planning.

Legends

Table 1. Distribution of images across curvature categories

Table 2. Performance of object detection models

Table 3. Segmentation model performance

Table 4. Comparison of geometric curvature parameters among canal severity groups

Table 5. Performance of classification models

Table 6. Performance of classification models across each curvature severity category

Supplementary Table S1. Pairwise post-hoc statistical comparisons (mean differences, 95% confidence intervals, and p -values) of the extracted geometric parameters across the straight, moderate, and severe root canal curvature classes.

Supplementary Table S2 Comprehensive 10-fold conditional cross-validation performance metrics for the ablation study, comparing the isolated geometric, isolated CNN (ResNet-50), and hybrid fusion classification models.

Figure 1: Illustration of Lütten's angle and Pruett's radius where AB represents the coronal central axis and CD represents the apical central axis of the canal. The acute angle formed at the intersection of AB and CD is denoted as α (Luiten's angle). R is the radius of the circle at the point of maximum curvature, corresponding to Pruett's radius.

Figure 2: Illustration of canal access, curvature height and curvature distance. AB represents the line connecting the canal orifice and the canal terminus, while AC is a line drawn parallel to the long axis of the canal from the coronal aspect. D represents the point of maximum curvature. The angle α denotes the canal access angle, X represents the curvature height, and Y represents the curvature distance.

Figure 3: Automated geometric assessment of a straight root canal. a) Preoperative radiograph, (b) RF-DETR-based object detection with cropped premolar region of interest (ROI), (c) segmentation of the root canal space using YOLOv11, (d) Luiten's method demonstrating the internal angle between the coronal and apical axes, (e) Pruett's method estimating the radius of curvature using a best-fit circle, and (f) Canal Access Angle (CAA) analysis, curvature height and distance.

Figure 4: Automated geometric assessment of a moderately curved root canal. (a) Preoperative radiograph, (b) RF-DETR-based object detection with cropped premolar region of interest (ROI), (c) segmentation of the root canal space using YOLOv11, (d) Luiten's method demonstrating the internal angle between the coronal and apical axes, (e) Pruett's method estimating the radius of curvature using a best-fit circle, and (f) Canal Access Angle (CAA) analysis, curvature height and distance.

Figure 5: Automated geometric assessment of a severely curved root canal. a) Preoperative radiograph, (b) RF-DETR-based detection with cropped premolar region of interest (ROI), (c) segmentation of the root canal space using YOLOv11, (d) Luiten's method demonstrating the internal angle between the coronal and apical axes, (e) Pruett's method estimating the radius of curvature using a best-fit circle, and (f) Canal Access Angle (CAA) analysis, curvature height and distance.

Figure 6: Morphological distribution of geometric parameters across root canal severity classes. The violin plots visualize the data density, median, and interquartile ranges for: Luiten's Angle, Pruett's Radius, Canal Access Angle, Curvature Height, and Curvature Distance. The distributions visually demonstrate a proportional spike in Luiten's Angle, Canal Access Angle, and Curvature Height, alongside a significant contraction in Pruett's Radius, as morphological severity increases, confirming the clinical validity of the extracted features.

Figure 7: Comparative diagnostic performance of the ablation framework evaluated via stratified 10-fold cross-validation. The left column panels display the Receiver Operating Characteristic (ROC) curves for the (a) Geometric-only model, (c) CNN-only (ResNet-50) model, and (e) Hybrid Fusion model across the straight, moderate, and severe curvature classes. The right column panels present the corresponding Precision-Recall (PR) curves for the (b) Geometric-only model, (d) CNN-only model, and (f) Hybrid Fusion model.

Figure 8: XGBoost feature importance rankings across the ablation framework. (a) Isolated geometric model, (b) Isolated CNN model (ResNet-50), and (c) Hybrid feature-fusion model, illustrating the contribution of geometric and deep learning features to curvature severity classification.

Figure 9: Confusion matrices detailing the class-specific predictive distributions for the (a) Geometric-only, (b) CNN-only, and (c) Hybrid Fusion models.

Supplementary Figure S1: Schematic overview of the experimental multi-model

computational pipeline. The automated framework consists of (1) Input, (2) premolar detection, (3) root canal segmentation, (4) multiparametric feature extraction combining geometric and deep learning parameters, and (5) final root canal curvature classification.

Supplementary Figure S2: Data flow diagram detailing the four-stage computational pipeline evaluated via 10-fold stratified cross-validation on 1,644 periapical radiographs.

REFERENCES

- Albitar, L., Zhao, T., Huang, C., & Mahdian, M. (2022). Artificial Intelligence (AI) for Detection and Localization of Unobturator Second Mesial Buccal (MB2) Canals in Cone-Beam Computed Tomography (CBCT). *Diagnostics (Basel, Switzerland)*, **12**
- AL Qurri, A., & Almekkawy, M. (2023). Improved UNet with Attention for Medical Image Segmentation. *Sensors*, **23**(20), 8589.
- American Association of Endodontists. (2018). AAE Endodontic Case Difficulty Assessment Form and Guidelines. *American Association of Endodontists*. (accessed on 23 March 2026)
- Arias, A., & Peters, O. A. (2022). Present status and future directions: Canal shaping. *International Endodontic Journal*, **55** Suppl 3(Suppl 3), 637–655.
- Bui, R., Iozzino, R., Richert, R., Roy, P., Boussel, L., Tafrount, C., & Ducret, M. (2023). Artificial Intelligence as a Decision-Making Tool in Forensic Dentistry: A Pilot Study with I3M. *International Journal of Environmental Research and Public Health*, **20**(5), 4620.
- Chaniotis, A., & Ordinola-Zapata, R. (2022). Present status and future directions: Management of curved and calcified root canals. *International Endodontic Journal*, **55** Suppl 3, 656–684.
- Chen, Z., Liu, Q., Wang, J., Ji, N., Gong, Y. & Gao, B. (2025) Tooth image segmentation and root canal measurement based on deep learning. *Frontiers in Bioengineering and Biotechnology* **13**, 1565403.
- Duman, Ş. B., Çelik Özen, D., Bayrakdar, I. Ş., Baydar, O., Alhaija, E. S. A., Helvacioğlu Yiğit, D., Çelik, Ö., Jagtap, R., Pileggi, R., & Orhan, K. (2024). Second mesiobuccal canal segmentation with YOLOv5 architecture using cone beam computed tomography images. *Odontology*, **112**(2), 552–561.

- Essam, O., Boyle, E. L., Whitworth, J. M., & Jarad, F. D. (2021). The Endodontic Complexity Assessment Tool (E-CAT): A digital form for assessing root canal treatment case difficulty. *International Endodontic Journal*, **54**(7), 1189–1199.
- Günday, M., Sazak, H., & Garip, Y. (2005). A comparative study of three different root canal curvature measurement techniques and measuring the canal access angle in curved canals. *Journal of Endodontics*, **31**(11), 796–798.
- Hilmi, A., Patel, S., Mirza, K., & Galicia, J. C. (2023). Efficacy of imaging techniques for the diagnosis of apical periodontitis: A systematic review. *International Endodontic Journal*, **56** Suppl 3, 326–339.
- Hülsmann, M., Peters, O. A., & Dummer, P. M. H. (2005). Mechanical preparation of root canals: Shaping goals, techniques, and means. *Endodontic Topics*, **10**(1), 30–76.
- Karkehabadi, H., Khoshbin, E., Ghasemi, N., Mahavi, A., Mohammad-Rahimi, H., & Sadr, S. (2024). Deep learning for determining the difficulty of endodontic treatment: a pilot study. *BMC Oral Health*, **24**(1), 574.
- Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, J., Prasadha, M. K., Pei, J., Ting, M. Y. L., Zhu, J., Li, C., Hewett, S., Dong, J., Ziyar, I., ... Zhang, K. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*, **172**(5), 1122–1131.e9.
- Khalifa, M., & Albadawy, M. (2024). AI in diagnostic imaging: Revolutionising accuracy and efficiency. *Computer Methods and Programs in Biomedicine Update*, **5**, Article 100146.
- Kirkevang, L. L., Ørstavik, D., Hörsted-Bindslev, P., & Wenzel, A. (2000). Periapical status and quality of root fillings and coronal restorations in a Danish population. *International Endodontic Journal*, **33**(6), 509–515.
- Kazimierczak, W., Wajer, R., Wajer, A., Kalka, K., Kazimierczak, N. & Serafin, Z. (2024) Evaluating the diagnostic accuracy of an AI-driven platform for assessing endodontic treatment outcomes using panoramic radiographs: a preliminary study. *Journal of Clinical Medicine* **13**, 3401.
- Lahoud, P., EzEldeen, M., Beznik, T., Willems, H., Leite, A., Van Gerven, A., & Jacobs, R. (2021). Artificial Intelligence for Fast and Accurate 3-Dimensional Tooth Segmentation on Cone-beam Computed Tomography. *Journal of Endodontics*, **47**(5), 827–835.
- Lin, L. M., Rosenberg, P. A., & Lin, J. (2005). Do procedural errors cause endodontic treatment

- failure? *Journal of the American Dental Association* (1939), **136**(2), 187–193; quiz 231.
- Luiten, D. J., Morgan, L. A., Baugartner, J. C., & Marshall, J. G. (1995). A comparison of four instrumentation techniques on apical canal transportation. *Journal of Endodontics*, **21**(1), 26–32.
- Mohammad-Rahimi, H., Sohrabniya, F., Ourang, S. A., Dianat, O., Aminoshariae, A., Nagendrababu, V., Dummer, P. M. H., Duncan, H. F., & Nosrat, A. (2024). Artificial intelligence in endodontics: Data preparation, clinical applications, ethical considerations, limitations, and future directions. *International Endodontic Journal*, **57**(11), 1566–1595.
- Nagendrababu, V., Pigg, M., Duncan, H.F., Abbott, P.V., Fouad, A.F., Kruse, C., Patel, S., Rechenberg, D.K., Setzer, F.C., Rossi-Fedele, G. & Dummer, P.M.H. (2024) PRIDASE 2024 guidelines for reporting diagnostic accuracy studies in endodontics: A consensus-based development. *International Endodontic Journal*. 2024 Aug;57(8):996-1005
- Ng, Y. L., Mann, V., & Gulabivala, K. (2011). A prospective study of the factors affecting outcomes of nonsurgical root canal treatment: Part 1: Periapical health. *International Endodontic Journal*, **44**(7), 583–609.
- Ourang, S. A., Sohrabniya, F., Sadr, S., Lee, J., Ramzy, N., Law, A. S., Lam, E. W. N., Khan, A. A., & Nosrat, A. (2025). Artificial Intelligence in the Detection of Clinically Negotiable Second Mesio-Buccal Canals in Periapical Images of Maxillary Molars. *International Endodontic Journal*. <https://doi.org/10.1111/iej.70065>
- Patel, S., Dawood, A., Ford, T. P., & Whaites, E. (2007). The potential applications of cone beam computed tomography in the management of endodontic problems. *International Endodontic Journal*, **40**(10), 818–830.
- Patel, S., Brown, J., Semper, M., Abella, F., & Mannocci, F. (2019). European Society of Endodontology position statement: Use of cone beam computed tomography in Endodontics. *International Endodontic Journal*, **52**(12), 1675–1678.
- Pinto, J. C., de Faria Vasconcelos, K., Leite, A. F., Wanderley, V. A., Pauwels, R., Oliveira, M. L., Jacobs, R., & Tanomaru-Filho, M. (2023). Image quality for visualization of cracks and fine endodontic structures using 10 CBCT devices with various scanning protocols and artefact conditions. *Scientific Reports*, **13**(1), 4001.
- Pruett, J. P., Clement, D. J., & Carnes, D. L., Jr (1997). Cyclic fatigue testing of nickel-titanium

- endodontic instruments. *Journal of endodontics*, **23**(2), 77–85.
- Qiao, X., Xu, T., Chen, L., & Yang, D. (2021). Analysis of Root Canal Curvature and Root Canal Morphology of Maxillary Posterior Teeth in Guizhou, China. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, **27**, e928758.
- Rashid, H., Balani, P., & Niazi, F. (2015). A brief review of the methods used to determine the curvature of root canals. *Journal of Restorative Dentistry*, **3**(3), 57.
- Schneider, S. W. (1971). A comparison of canal preparations in straight and curved root canals. *Oral Surgery, Oral Medicine, Oral Pathology*, **32**, 271–275.
- Sonntag, D., Stachniss-Carp, S., Stachniss, C., & Stachniss, V. (2006). Determination of root canal curvatures before and after canal preparation (part II): A method based on numeric calculus. *Australian Endodontic Journal: The Journal of the Australian Society of Endontology Inc*, **32**(1), 16–25.
- Sousa Melo, S. L., Fayad, M. I., Gohel, A., Johnson, B. R., Kalathingal, S., Mahdian, M., Nair, M., Setzer, F. C., & Makins, S. R. (2026). AAE and AAOMR Joint Position Statement: Use of Cone-Beam Computed Tomography in Endodontics 2025 Update. *Journal of Endodontics*, **52**(1), 4–13.
- Special Committee to Revise the Joint AAE/AAOMR Position Statement on use of CBCT in Endodontics. (2015). AAE and AAOMR Joint Position Statement: Use of Cone Beam Computed Tomography in Endodontics 2015 Update. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, **120**(4), 508–512.
- Schwendicke, F., Samek, W. & Krois, J. (2020) Artificial intelligence in dentistry: chances and challenges. *Journal of Dental Research* **99**, 769–774.
- Slim, M.L., Jacobs, R., de Souza Leal, R.M. & Fontenele, R.C. (2024) AI-driven segmentation of the pulp cavity system in mandibular molars on CBCT images using convolutional neural networks. *Clinical Oral Investigations* **28**, 650.
- Tejani, A. S., Klontzas, M. E., Gatti, A. A., Mongan, J. T., Moy, L., Park, S. H., Kahn, C. E., Jr, & CLAIM 2024 Update Panel (2024). Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update. *Radiology. Artificial intelligence*, **6**(4), e240300.
- Thanaruengrong, P., Kulvitit, S., Navachinda, M., & Charoenlarp, P. (2021). Prevalence of complex root canal morphology in the mandibular first and second premolars in Thai

- population: CBCT analysis. *BMC Oral Health*, **21**(1), 449. Weine, F. S. (1982). *Endodontic therapy* (3rd ed., pp. 256–340). St. Louis, MO: C. V. Mosby.
- Wang, Y., Li, Z., Wu, C., Liu, J., Zhang, Y., Chen, J. et al. (2025) MICCAI STSR 2025 Challenge: Semi-Supervised Teeth and Pulp Segmentation and CBCT-IOS Registration. arXiv [Preprint]. Available from: <https://arxiv.org/abs/2512.02867>.
- Zhu, Y., Gu, Y., Du, R., & Li, C. (2003). Reliability of two methods on measuring root canal curvature. *International Chinese Journal of Dentistry*, **3**, 118–121.

Table 1. Distribution of images across curvature categories

Root Canal Curvature	Total Images (n)	Percentage of Dataset (%)	Training Set per Fold (~90%)	Test Set per Fold (~10%)
Straight	633	38.50	570	63
Moderate	885	53.80	796	89
Severe	126	7.70	113	13
Total	1,644	100	~1,479	~165

Table 2. Performance of object detection models

Model	Precision (%)	Recall (%)	F1-score (%)	<u>mAP@0.5 (%)</u>	<u>mAP@0.5-0.95 (%)</u>
YOLOv8	93.5 ± 1.2	89.5 ± 0.4	91.5 ± 1.1	98.3 ± 1.4	91.4 ± 1.3
YOLOv11	91.6 ± 1.1	93.8 ± 1.3	92.7 ± 1.0	93.9 ± 0.9	86.2 ± 1.0
RF-DETR	93.3 ± 0.2	97.5 ± 0.8	95.4 ± 0.4	97.6 ± 0.7	84.8 ± 1.2

Table 3. Segmentation model performance

Model	Dice (%)	IoU (%)	Precision (%)	Recall (%)	clDice (%)	HD95 (mm)	ASSD (mm)
Attention U-Net	86.17	76.72	88.21	84.97	89.08	0.8457	0.3291
YOLOv8-seg	83.59	76.37	88.83	83.78	88.33	0.8591	0.1992
YOLOv11-seg	85.87	77.53	89.24	87.57	90.36	0.5813	0.159

Table 4. Comparison of geometric curvature parameters among canal groups

Parameter	Straight	Moderate	Severe	F(2,1641)	p	η^2
Lüiten's Angle (°)	4.53 ± 3.57	13.61 ± 5.73	32.57 ± 15.66	1098.13	<0.001	0.572
Pruett's Radius (mm)	44.94 ± 43.27	16.62 ± 13.79	7.59 ± 6.51	209.53	<0.001	0.203
Canal Access Angle (°)	2.03 ± 1.86	4.22 ± 3.00	8.93 ± 6.76	281.17	<0.001	0.255
Curvature Height (mm)	0.13 ± 0.07	0.32 ± 0.15	0.61 ± 0.35	623.82	<0.001	0.432
Curvature Distance (mm)	4.39 ± 2.17	4.78 ± 1.66	4.72 ± 1.67	8.08	<0.001	0.010

Post-hoc pairwise comparisons (Tukey's HSD) revealed statistically significant differences between the curvature groups for Lüiten's Angle, Pruett's Radius, Canal Access Angle, and Curvature Height ($p < 0.05$). For Curvature Distance, a significant difference was observed between the Straight and Moderate groups ($p < 0.001$), whereas differences between the Straight and Severe groups ($p = 0.183$) and between the Moderate and Severe groups ($p = 0.927$) were not significant.

Table 5. Performance of classification models

Model Feature Set	Accuracy (%)	Macro-Precision (%)	Macro-Recall (%)	Macro F1-Score (%)
Geometric Only	88.38± 1.77	87.28 ± 4.80	86.30 ± 5.67	86.61± 5.00
CNN Only (ResNet-50)	82.42± 2.19	82.50 ± 5.01	70.49 ± 5.58	73.87 ± 5.48
Hybrid Fusion	89.60± 2.11	91.63 ± 2.89	83.80 ± 5.89	86.77 ± 4.66

Table 6. Performance of classification models across each curvature severity category

Model	Class	Precision (%)	Recall / Sensitivity (%)	F1-Score (%)
Geometric Only	Straight	91.12 $\pm$ 4.02	89.19 $\pm$ 3.61	90.10 $\pm$ 3.14
	Moderate	92.22 $\pm$ 1.94	89.08 $\pm$ 3.62	90.60 $\pm$ 2.60
	Severe	78.51 $\pm$ 10.56	80.62 $\pm$ 15.51	79.14 $\pm$ 11.68
CNN Only (ResNet-50)	Straight	87.24 $\pm$ 3.59	82.63 $\pm$ 6.20	84.69 $\pm$ 3.02
	Moderate	83.58 $\pm$ 3.67	88.26 $\pm$ 3.00	85.80 $\pm$ 2.38
	Severe	76.68 $\pm$ 13.39	40.58 $\pm$ 17.56	51.12 $\pm$ 16.95
Hybrid Fusion	Straight	93.08 $\pm$ 2.63	90.20 $\pm$ 2.79	91.57 $\pm$ 1.62
	Moderate	91.43 $\pm$ 2.89	92.09 $\pm$ 2.05	91.73 $\pm$ 1.86
	Severe	90.39 $\pm$ 7.50	69.10 $\pm$ 17.02	77.00 $\pm$ 11.72

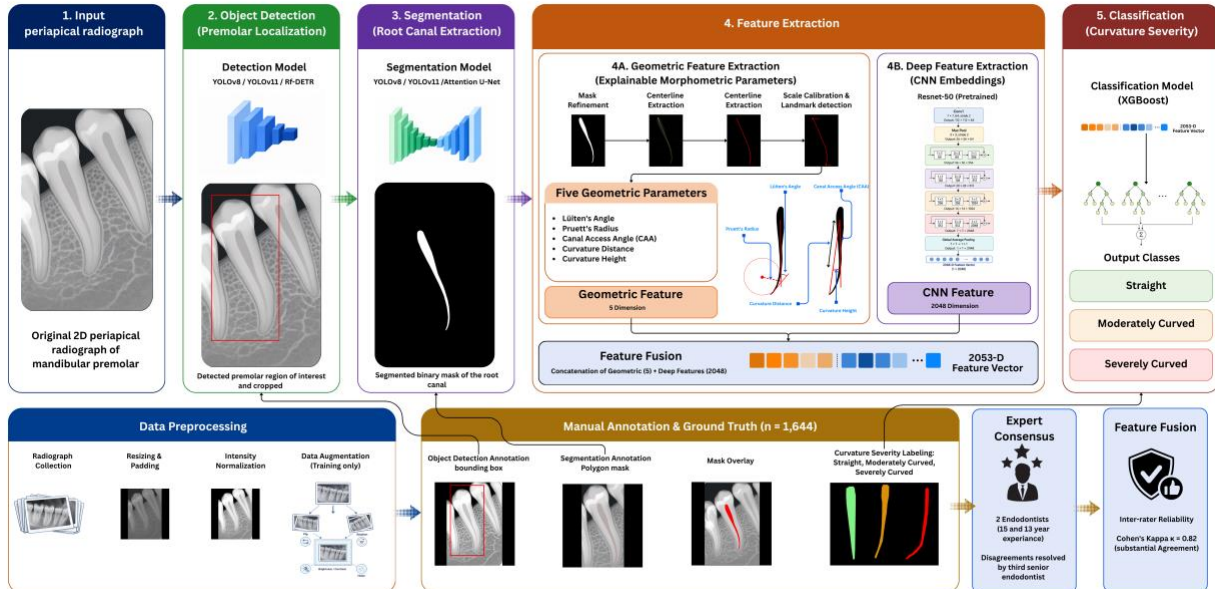
Supplementary table s1 Pairwise post-hoc statistical comparisons (mean differences, 95% confidence intervals, and *p*-values) of the extracted geometric parameters across the straight, moderate, and severe root canal curvature classes.

Parameter	Straight vs. Moderate	Straight vs. Severe	Moderate vs. Severe
Lüiten's Angle (°)	-9.08 [-9.86, -8.29]p<0.001	-28.04 [-29.51, -26.57]p<0.001	-18.97 [-20.40, -17.53]p<0.001
Pruett's Radius (mm)	28.32 [24.81, 31.83]p<0.001	37.35 [30.77, 43.93]p<0.001	9.03 [2.60, 15.45]p=0.003
Canal Access Angle (°)	-2.18 [-2.56, -1.80]p<0.001	-6.89 [-7.61, -6.18]p<0.001	-4.71 [-5.41, -4.02]p<0.001
Curvature Height (mm)	-0.19 [-0.21, -0.17]p<0.001	-0.48 [-0.51, -0.44]p<0.001	-0.29 [-0.32, -0.25]p<0.001
Curvature Distance (mm)	-0.39 [-0.62, -0.16]p<0.001	-0.32 [-0.75, 0.11]p=0.183 (n.s.)	0.07 [-0.35, 0.48]p=0.927 (n.s.)

Supplementary table s2 Comprehensive 10-fold conditional cross-validation performance metrics for the ablation study, comparing the isolated geometric, isolated CNN (ResNet-50), and hybrid fusion classification models.

	Geometric Mean $\pm$ SD	CNN Mean $\pm$ SD	Hybrid Mean $\pm$ SD
Accuracy (%)	90.63 $\pm$ 2.93	84.53 $\pm$ 2.19	91.89 $\pm$ 2.11
Balanced_accuracy (%)	88.83 $\pm$ 5.47	72.46 $\pm$ 5.05	86.36 $\pm$ 6.21
Cohen_kappa	0.832 $\pm$ 0.052	0.712 $\pm$ 0.041	0.852 $\pm$ 0.040
Mcc	0.833 $\pm$ 0.052	0.716 $\pm$ 0.040	0.853 $\pm$ 0.039
Gmean	88.46 $\pm$ 6.04	66.83 $\pm$ 11.12	85.17 $\pm$ 7.67
Macro_f1 (%)	87.89 $\pm$ 4.93	75.08 $\pm$ 5.43	88.18 $\pm$ 4.70
Weighted_f1 (%)	90.63 $\pm$ 2.95	83.81 $\pm$ 2.21	91.72 $\pm$ 2.25
Roc_auc_macro	0.967 $\pm$ 0.019	0.948 $\pm$ 0.011	0.980 $\pm$ 0.012
Roc_auc_weighted	0.966 $\pm$ 0.012	0.946 $\pm$ 0.011	0.981 $\pm$ 0.006
Pr_auc_macro	0.935 $\pm$ 0.039	0.869 $\pm$ 0.038	0.952 $\pm$ 0.030
Pr_auc_weighted	0.955 $\pm$ 0.018	0.925 $\pm$ 0.017	0.973 $\pm$ 0.013
Log_loss	0.290 $\pm$ 0.078	0.424 $\pm$ 0.054	0.247 $\pm$ 0.072
Brier_score_avg	0.049 $\pm$ 0.012	0.077 $\pm$ 0.010	0.042 $\pm$ 0.011

Supplementary Figure S1. Schematic overview of the experimental multi-model computational pipeline. The automated framework consists of (1) Input, (2) premolar detection, (3) root canal segmentation, (4) multiparametric feature extraction combining geometric and deep learning parameters, and (5) final root canal curvature classification



Supplementary Figure S2.

Data flow diagram detailing the four-stage computational pipeline evaluated via 10-fold stratified cross-validation on 1,644 periapical radiographs.

