Journal of Pharmaceutical and Medical Research
Print ISSN: 2663-1954
Online ISSN: 2663-1962
DOI: https://doi.org/10.61784/jpmr3091

DESIGN OF AN INTELLIGENT AUXILIARY GRADING SCHEME FOR DENTAL CARIES BASED ON DENTAL IMAGE SEGMENTATION AND FEATURE QUANTIFICATION

RongMi Mao[1*], ZhuoHeng Li[2]
[1]*School of Science, Shandong Jianzhu University, Jinan 250101, Shandong, China.*
[2]*School of Computer and Artificial Intelligence, Shandong Jianzhu University, Jinan 250101, Shandong, China.*
[*]*Corresponding Author: RongMi Mao*

Abstract: Dental caries screening in primary care is challenged by small early-stage lesions, blurred boundaries, variable image quality, and subjective interpretation. This paper proposes an intelligent auxiliary grading scheme based on TransUNet and feature quantification. After grayscale normalization, contrast-limited adaptive histogram equalization, Gaussian denoising, and quality assessment, the model combines convolutional local-feature extraction with Transformer-based global context modeling to generate pixel-level lesion probability maps and binary masks. Connected-region analysis then quantifies lesion area, location, contour, overlap, and regional confidence. A shared classification branch produces C0-C3 auxiliary grades, while probability entropy, mask morphology, image-quality flags, and lesion confidence jointly trigger clinician review. The scheme defines patient-level dataset separation, annotation cross-checking, comparative and ablation experiments, and auditable outputs linking each prediction to its source image, preprocessing parameters, model version, mask, features, grade probabilities, and review status. Dice, IoU, precision, recall, accuracy, and F1-score are specified as validation metrics. The resulting framework is interpretable and traceable, but remains an auxiliary screening design requiring clinical calibration and external validation.
Keywords: Dental caries; Image segmentation; TransUNet; Multi-task learning; Feature quantification; Clinician review

1 INTRODUCTION

Dental caries is a common hard-tissue disease encountered in primary oral screening. Detecting a lesion at an early stage can support timely intervention and reduce subsequent tissue damage. Bitewing, periapical, panoramic, cephalometric, cone-beam computed tomography, and intraoral images provide different views of tooth crowns, roots, proximal surfaces, and surrounding structures. Figure 1 retains the representative image categories presented in the source material.

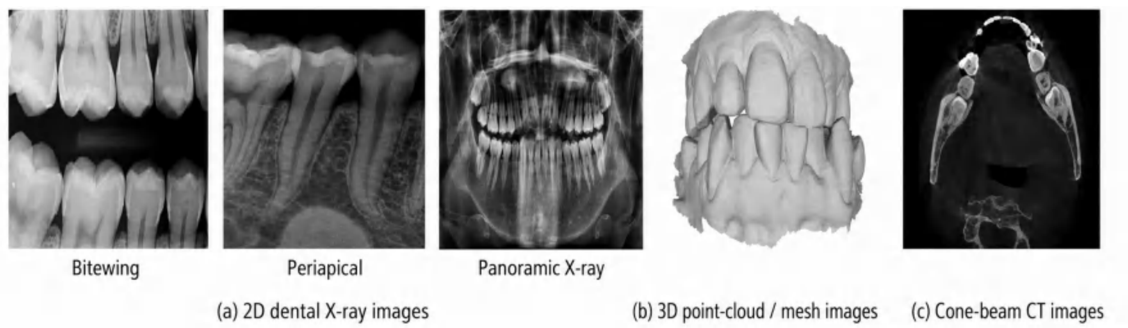

Figure 1 Representative Dental Image Types Used in Caries Screening

Early lesions may have subtle gray changes, limited spatial extent, blurred borders, and irregular morphology. Screening performance is consequently affected by exposure, sharpness, contrast, image completeness, and clinical experience. Image-level classification can indicate whether caries may be present but cannot show its exact position. Object detection places a box around a suspicious region, yet the box does not describe the true lesion boundary. Pixel-level segmentation is more suitable for extracting lesion area, location, contour, and confidence as interpretable evidence.

Deep learning has been applied to two-dimensional radiographs, three-dimensional oral data, and CBCT for tooth localization, numbering, segmentation, lesion detection, and computer-aided diagnosis [1-2]. Domestic research has explored dental-lesion detection, pediatric caries recognition, oral-disease imaging, maxillofacial image analysis, and panoramic tooth segmentation [3-7]. Recent reviews emphasize clinical translation, explainability, data quality, and

human oversight [8]. International studies confirm the potential of convolutional networks and U-Net variants, while identifying limited datasets, costly annotation, and insufficient generalization as continuing constraints [9-15].
The research objective is to construct a reproducible scheme that converts a dental image into a quality record, lesion probability map, binary mask, quantitative feature set, C0-C3 auxiliary grade, uncertainty score, and review recommendation. The design specifies data preparation, model architecture, mathematical definitions, verification criteria, and the boundary between automated analysis and clinician confirmation. No unreported clinical performance is assumed.

2 OVERALL SCHEME AND DATA PREPARATION

2.1 Scheme Architecture

The processing sequence consists of image input, quality assessment, standardized preprocessing, lesion segmentation, feature quantification, auxiliary grading, uncertainty evaluation, and structured output. Figure 2 retains the source workflow. Each stage generates evidence that can be reviewed independently: a quality flag, probability map, lesion mask, quantitative indicators, class probabilities, and a final review status.

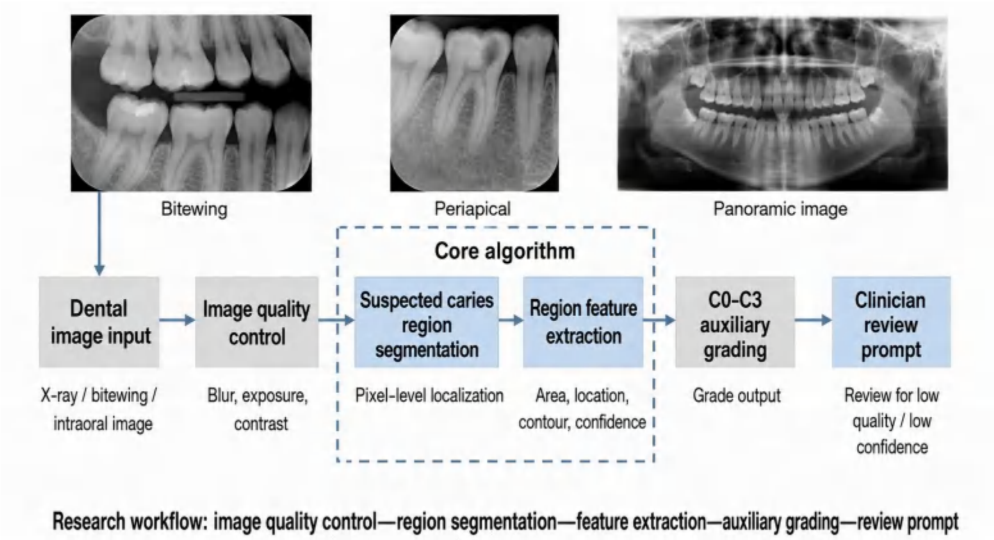

Figure 2 Overall Workflow for Segmentation-Driven Caries Grading

Table 1 Interpretation of the C0-C3 Auxiliary Grades

Grade	Auxiliary interpretation	Principal evidence
C0	No obvious carious lesion	No evident suspicious mask and low overall risk probability
C1	Suspected early lesion	Small, low-contrast region with a weak boundary
C2	Moderate-risk lesion	Clearer region with concerning area, position, or gray change
C3	High-risk lesion	Large lesion extent or a high model risk indication

The four grades are auxiliary categories rather than final diagnoses. C0 indicates no obvious lesion in the available image. C1 marks a small or weak suspicious region. C2 denotes a clearer lesion for which area, location, or gray change raises concern. C3 identifies a large lesion extent or a high-risk model output. Poor image quality, unstable prediction, abnormal mask morphology, or a C2-C3 output can trigger clinician review. Table 1 summarizes these auxiliary categories and their principal evidence.

2.2 Dataset Construction and Annotation

The data scheme prioritizes bitewing and periapical radiographs and accepts panoramic radiographs and intraoral images as complementary inputs. Existing lesion masks are retained directly. Image-level labels are converted into initial pixel annotations with Labelme, followed by two-person cross-checking and disagreement review. Low-quality images are not automatically discarded because difficult samples are needed for robustness assessment and review-trigger design.
Figure 3 records the source sequence from 713 raw images to 1,517 bitewing images and 6,032 single-tooth samples. Patient-level separation is required when constructing training, validation, and test sets so that images from one person do not appear in more than one subset. Annotation records retain lesion boundaries, image origin, quality status, and disagreement-resolution information.

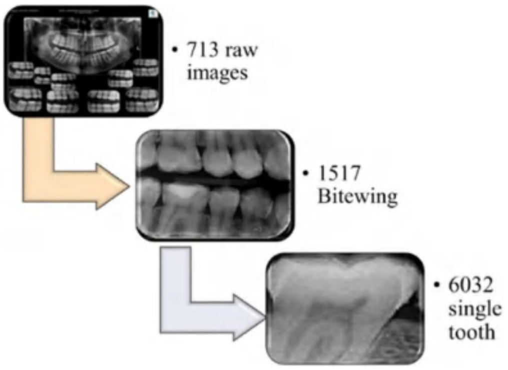

Figure 3 Source Dataset Construction and Sample Conversion

2.3 Standardized Preprocessing and Quality Control

Gray normalization reduces brightness differences between acquisition sources.

$$I_{\text{norm}}(x,y) = \frac{I(x,y)-I_{\min}}{I_{\max}-I_{\min}} \tag{1}$$

Contrast-limited adaptive histogram equalization enhances low-contrast lesions while controlling local noise amplification. Its local gray transformation is defined by the cumulative histogram density.

$$T(r) = r_{\min} + (r_{\max} - r_{\min}) \int_0^r p(t)dt \tag{2}$$

Gaussian smoothing suppresses random noise while retaining lesion boundaries. The source scheme sets the Gaussian standard deviation to 1.0. The corresponding image-quality control decisions are summarized in Table 2.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{3}$$

$$I_{\text{smooth}} = I * G \tag{4}$$

Table 2 Image-Quality Control Decisions

Quality condition	Processing decision	Retained record
Adequate clarity, brightness, contrast, and completeness	Enter standardized preprocessing	Quality score
Blur, underexposure, or low contrast	Enhance and mark for review	Image and quality flags
Severe incompleteness or unusable acquisition	Recommend retake	Retake reason
Difficult but interpretable sample	Retain for robustness analysis	Review status

3 TRANSUNET-BASED MULTI-TASK MODEL

3.1 Encoder and Global Context Modeling

The model combines a convolutional encoder, Transformer encoder, U-Net decoder, and severity-classification branch. Figure 4 retains the proposed structure. The convolutional encoder extracts local texture, edge, and gray-change features at 256 x 256, 128 x 128, 64 x 64, and 32 x 32 resolutions. The highest-level feature map is flattened into 1,024 tokens with 512 channels and processed by eight attention heads.

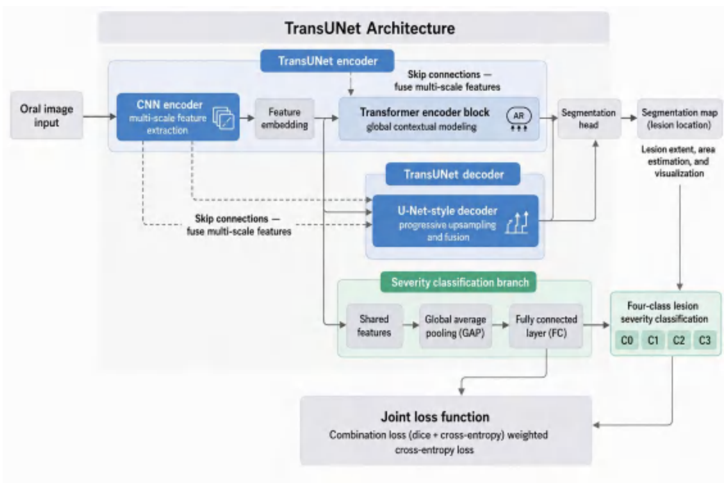

Figure 4 Proposed TransUNet Multi-Task Architecture

$$F_{\text{cnn}}{}^{i} = \text{Conv}(\text{BN}(\text{ReLU}(F_{\text{cnn}}{}^{i-1}))) \tag{5}$$

Self-attention relates distant tooth regions so that a local gray variation can be interpreted against overall anatomy. Query, key, and value projections are combined by scaled dot-product attention. With eight heads and 512 channels, the dimension assigned to each head is 64.

$$\text{Attention}(Q, K, V) = \text{Softmax}(\frac{QK^T}{d_k^{\frac{1}{2}}})V \tag{6}$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{Attn}_1, \ldots, \text{Attn}_h)W^O \tag{7}$$

The source further uses local-window attention to reduce unnecessary interaction outside the relevant tooth region. The feature map is partitioned into windows, attention is calculated inside each window, and shifted windows exchange information across neighboring boundaries. Figure 5 retains this contextual modeling mechanism. It complements the convolutional encoder by linking separated parts of the same tooth and neighboring structures while preserving computational feasibility.

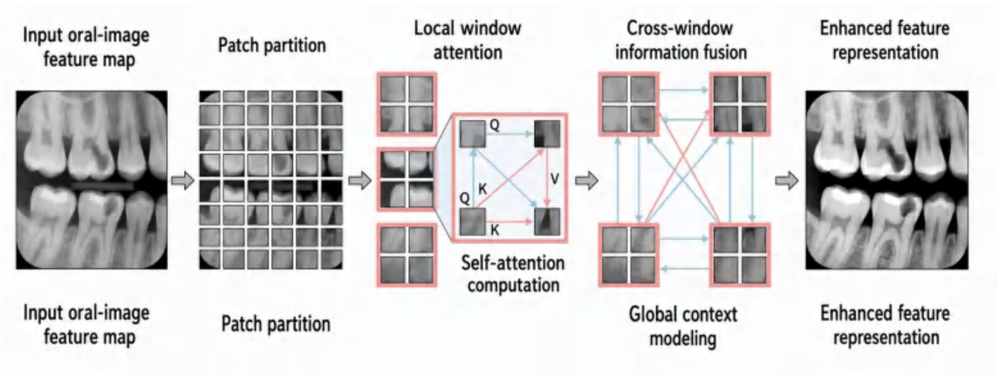

Figure 5 Local-Window Transformer Attention Retained from the Source Design

3.2 Decoder, Segmentation Output, and Classification Branch

The decoder performs four stages of upsampling. Transposed convolution with a stride of two enlarges the feature map, and skip connections fuse high-level semantic information with low-level boundary details. This design addresses the weak borders and small spatial extent of early lesions.

$$F_{\text{up}}{}^{i} = \text{UpConv}(F_{\text{up}}{}^{i+1}) + F_{\text{cnn}}{}^{3-i} \tag{8}$$

$$M = \text{Sigmoid}(\text{Conv}(F_{\text{up}}{}^{0})) \tag{9}$$

The shared Transformer feature is also pooled and passed to a classification layer. The resulting probability vector contains the probabilities of C0, C1, C2, and C3, and the largest probability defines the auxiliary grade.

$$P = \text{Softmax}(Wg + b) \tag{10}$$

$$y_{\text{cls}} = \text{argmax}(P) \tag{11}$$

3.3 Joint Optimization

Segmentation and grading are trained jointly. The segmentation objective combines Dice loss and binary cross-entropy, whereas the classification branch uses the grade label. The weights alpha and beta balance the spatial and categorical tasks.

$$L_{\text{total}} = \alpha L_{\text{seg}} + \beta L_{\text{cls}} \tag{12}$$

$$L_{\text{dice}} = 1 - \frac{2\sum M M_{\text{gt}} + \varepsilon}{\sum M + \sum M_{\text{gt}} + \varepsilon} \tag{13}$$

$$L_{\text{bce}} = -\frac{1}{N}\sum[M_{\text{gt}}\log M + (1 - M_{\text{gt}})\log(1 - M)] \tag{14}$$

The joint objective encourages the encoder to retain both boundary evidence and severity-related context. Segmentation gives the grade a visible spatial basis, while the grade branch provides global semantic supervision to the shared representation.

The two tasks use one image input and a shared encoder but produce different outputs. The segmentation branch preserves pixel-level evidence, and the classification branch summarizes the overall severity. During training, errors from both branches are propagated to the shared representation. This arrangement avoids the information limitation of a segmentation-only model and the weak spatial explanation of a classification-only model. Figure 6 illustrates the shared-encoder dual-task relationship between lesion segmentation and auxiliary grading.

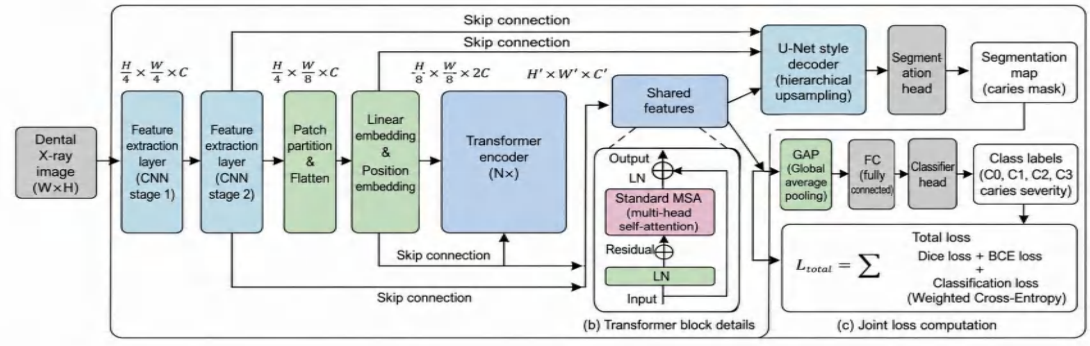

Figure 6 Dual-Task Principle for Lesion Segmentation and Auxiliary Grading

4 LESION QUANTIFICATION AND AUXILIARY GRADING

4.1 Probability Map, Binary Mask, and Connected Regions

The sigmoid output is retained as a lesion probability map and converted to a binary mask with the source threshold of 0.5. Connected-component analysis separates individual suspicious regions. Figure 7 preserves the source examples of lesion annotation and overlay output.

$$M_{\text{bin}} = \begin{cases} 1 & \text{if} M \geq T \\ 0 & \text{otherwise} \end{cases}, T = 0.5 \tag{15}$$

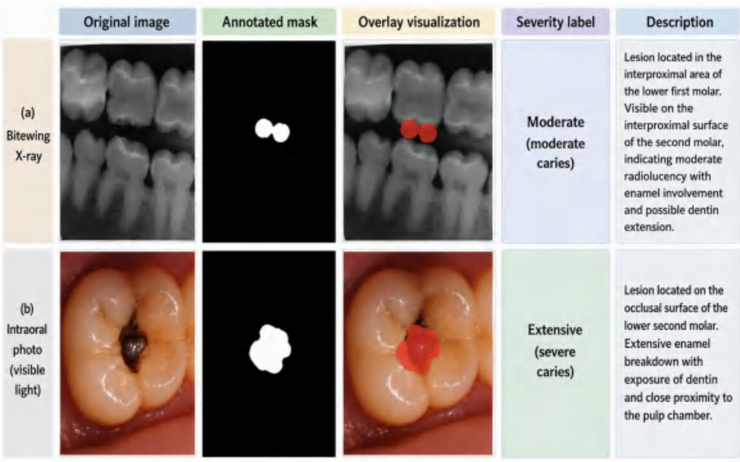

Figure 7 Lesion Annotation and Segmentation-Overlay Examples

4.2 Quantitative Features

For each connected region, the scheme extracts pixel area, physical area, centroid, tooth-relative position, contour coordinates, perimeter, overlap with the reference mask, and average regional confidence. If pixel spacing is available, physical area is obtained by multiplying the pixel area by the squared spacing. The source proposes two adjustable area references, 2.0 mm2 and 5.0 mm2, to support grade analysis; they remain subject to annotation statistics and expert review. The resulting quantitative outputs and their interpretive roles are summarized in Table 3.

$$S_i = \text{Area}(C_i) \times \text{PixelSpacing}^2 \tag{16}$$

$$\text{IoU} = \frac{\text{Area}(M_{\text{bin}} \cap M_{\text{gt}})}{\text{Area}(M_{\text{bin}} \cup M_{\text{gt}})} \tag{17}$$

$$\text{Confidence}(C_i) = \frac{\sum_{(x,y)\in C_i} M(x,y)}{\text{Area}(C_i)} \tag{18}$$

Table 3 Quantitative Outputs Derived from the Lesion Mask

Output	Definition	Role in interpretation
Area	Pixel count or physical area of each connected region	Describes lesion extent
Location	Centroid and tooth-relative position	Identifies the suspicious tooth region
Contour	Boundary coordinates, perimeter, and shape	Supports visual verification
IoU	Overlap between predicted and reference masks	Evaluates spatial agreement
Confidence	Mean probability inside a connected region	Measures regional reliability

4.3 Uncertainty and Review Trigger

Classification uncertainty is measured by probability entropy. A uniform probability distribution produces higher uncertainty than a concentrated distribution. Review is triggered when image quality is poor, lesion confidence is low, entropy is high, the mask is tiny, fragmented, or morphologically abnormal, or the predicted grade is C2 or C3. The corresponding clinician-review trigger logic is summarized in Table 4.

$$U = -\sum_{k=0}^{3} P_k \log P_k \tag{19}$$

Table 4 Clinician-Review Trigger Logic

Evidence source	Trigger condition	Output action
Image quality	Blur, underexposure, low contrast, or incomplete field	Review or retake recommendation
Regional probability	Low mean confidence in a suspicious region	Manual boundary inspection
Class probability	High entropy across C0-C3	Manual grade confirmation
Mask morphology	Tiny, fragmented, or abnormal connected regions	Mask correction or rejection
Auxiliary grade	C2 or C3 output	Priority clinician review

5 VALIDATION DESIGN

5.1 Comparative and Ablation Experiments

The proposed TransUNet scheme is compared with U-Net and Attention U-Net under the same patient-level data split, input resolution, and annotation protocol. Segmentation is assessed with Dice, IoU, precision, and recall. Auxiliary grading is assessed with accuracy, recall, and F1-score. Evaluation is reported separately for image quality, lesion size, and grade so that aggregate performance does not conceal difficult cases. The planned model comparisons and ablation experiments are summarized in Table 5.

Table 5 Planned Validation Matrix

Experiment	Compared configuration	Primary evaluation purpose
Model comparison	U-Net, Attention U-Net, and TransUNet	Assess local and global feature modeling
Quality-control ablation	With and without the quality module	Assess robustness to acquisition defects
Quantification ablation	With and without regional features	Assess interpretability contribution
Multi-task ablation	Segmentation only versus segmentation plus grading	Assess task complementarity
Review ablation	Direct output versus review-trigger workflow	Assess risk-control behavior

The source design does not provide a completed experimental result table. Accordingly, the validation matrix defines what must be measured without fabricating numerical outcomes. Error analysis records false positive regions caused by restorations, overlap, or noise; false negative regions caused by weak contrast or small lesions; and grade disagreements associated with image quality or annotation ambiguity.

Training records include the input source, patient partition, preprocessing configuration, augmentation state, annotation version, loss components, and stopping criterion. Validation samples are used for model selection, while the test set remains isolated until the final comparison. The same thresholding and post-processing rules are applied to all models. These controls make the comparison attributable to model design rather than inconsistent data preparation.

5.2 Reproducibility and Output Audit

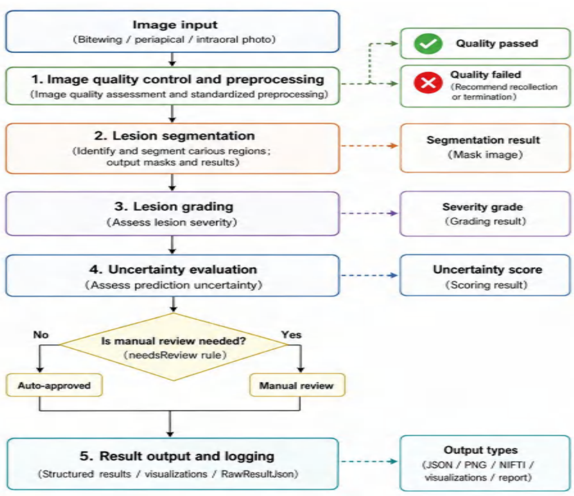

Figure 8 Technical Route of the Proposed Intelligent Grading Scheme

Each output is linked to its source image, preprocessing parameters, model version, probability map, binary mask, feature record, grade probabilities, and review status. The audit record allows a clinician to verify how the result was produced and to correct the mask or grade. The scheme is therefore evaluated not only by prediction metrics but also by traceability, interpretability, and the completeness of the review pathway.

The technical route in Figure 8 links data governance, model training, regional quantification, auxiliary grading, evaluation, and application-oriented output. The route also makes the human-review boundary explicit: difficult or uncertain samples are not forced into a definitive automated result.

6 DISCUSSION AND CONCLUSION

This scheme proposes an intelligent auxiliary grading system for dental caries based on TransUNet segmentation and feature quantification, with three technical characteristics: pixel-level segmentation provides visible lesion evidence; multi-task learning associates localization with C0-C3 grading; and a review trigger integrates image quality, confidence, entropy, morphology, and severity to prevent model outputs from serving as direct diagnostic decisions.

Each stage of the scheme (preprocessing, segmentation, quantification, grading, and review) has clearly defined inputs and outputs with traceable records, and the modular structure facilitates independent verification and component replacement. However, the system currently remains a research prototype, and its clinical feasibility has not yet been established through completed experiments.

Primary limitations include heterogeneous acquisition conditions across multiple imaging sources, high annotation costs with observer disagreement, difficulty in distinguishing early lesions from normal gray variation, and the need for clinical calibration of area thresholds and probability outputs. Future work should prioritize multi-center external validation, dataset expansion with patient-level stratification, stratified evaluation by image type and lesion scale, completion of comparative and ablation experiments, probability calibration, and prospective assessment of the human-machine review workflow. The system should always remain an auxiliary screening and review-prioritization tool, rather than a substitute for clinical diagnosis.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Zhao Y, Li J C, Cheng B D, et al. Applications and challenges of deep learning in oral medical imaging. Journal of Image and Graphics, 2024, 29(3): 586-607.
[2] She Y Y, Chen J Y, Gao F, et al. Research progress of deep learning in oral and maxillofacial imaging diagnosis. Chinese Journal of Stomatological Research (Electronic Edition), 2021, 15(3): 185-188.
[3] Liu F, Han M, Wan J, et al. Automatic detection algorithm for dental lesions based on deep learning. Chinese Journal of Lasers, 2022, 49(20): 2007207.
[4] Li R Z, Zhu J X, Wang Y Y, et al. Development of a prototype artificial-intelligence recognition system for pediatric caries based on deep learning. Chinese Journal of Stomatology, 2021, 56(12): 1253-1260.
[5] Wang Y L, Li G. Research progress of artificial intelligence in imaging diagnosis of oral diseases. Journal of Prevention and Treatment for Stomatological Diseases, 2022, 30(11): 816-820.
[6] Liu M Q, Fu K Y. Current status and prospects of deep-learning-assisted oral and maxillofacial medical imaging. Chinese Journal of Stomatology, 2023, 58(6): 533-539.
[7] Kou D Z. Automatic tooth segmentation method for panoramic radiographs based on deep learning. Frontiers of Data and Computing, 2024, 6(3).
[8] Dong X L. Innovation and clinical application of artificial intelligence in dentistry. Journal of Oral Science Research, 2025.
[9] Schwendicke F, Golla T, Dreher M, et al. Convolutional neural networks for dental image diagnostics: a scoping review. Journal of Dentistry, 2019, 91: 103226.
[10] Mohammad-Rahimi H, Motamedian S R, Rohban M H, et al. Deep learning for caries detection: a systematic review. Journal of Dentistry, 2022, 122: 104115.
[11] Lee S, OH S I, JO J, et al. Deep learning for early dental caries detection in bitewing radiographs. Scientific Reports, 2021, 11: 16807.
[12] Estai M, Mehdizadeh M, Vignarajan J, et al. Evaluation of a deep learning system for automatic detection of proximal surface dental caries on bitewing radiographs. Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology, 2022, 134(2): 262-270.
[13] Ying S, Huang F, Shen X, et al. Caries segmentation on tooth X-ray images with a deep network. Journal of Dentistry, 2022, 119: 104078.
[14] Baydar O, Rozylo-Kalinowska I, Peker I, et al. The U-Net approaches to evaluation of dental bite-wing radiographs: an artificial intelligence study. Diagnostics, 2023, 13(3): 453.
[15] Forouzeshfar P, Safaei A A, Ghaderi F, et al. Dental caries diagnosis from bitewing images using convolutional neural networks. BMC Oral Health, 2024, 24: 211.