The Bioscan AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES 21(1) 2322-2358, 2026 www.thebioscan.com

# AN INTELLIGENT DENTAL CYST PREDICTION FRAMEWORK USING PYRAMID SWIN VISION TRANSFORMER WITH REINFORCEMENT LEARNING - DRIVEN DIAGNOSTIC OPTIMIZATION

**Dr D Karthiga Rani**[1]
Assistant Professor
Department of BCA
RNS First Grade College, Bangalore, Karnataka, India
Email: karthigakumaresh@yahoo.com
**Dr C Radhamani**[2]
Associate Professor
Department of Mathematics
Kongu Arts and Science College (Autonomous), Erode, Tamil Nadu, India
Email: palaniradhu@gmail.com
**Dr S Prasath**[3]
Assistant Professor and Head
Department of Computer Science and Applications
VET Institute of Arts and Science (Co-education) College, Thindal, Erode, Tamil Nadu, India
Email: softprasaths@gmail.com

DOI: https://doi.org/10.63001/tbs.2026.v21.i01.pp2322-2358

**KEYWORDS**

*Dental Cyst Prediction,*
*Pyramid Swin Vision Transformer,*
*Reinforcement Learning,*
*Dental Radiographic Imaging,*
*Intelligent Diagnosis System,*
*Deep Learning,*
*Automated Lesion Detection,*
*Adaptive Diagnostic Optimization,*
*Clinical Decision Support System*

**Received on:29-02-2026**

**Accepted on: 15-03-2026**

**Published on: 26-03-2026**

## Abstract

Radiographic imaging is essential for the early and accurate prediction of dental cysts to prevent jaw destruction, tooth displacement, and delayed clinical treatment. Existing diagnostic methods rely heavily on expert interpretation and handcrafted features, resulting in inter-observer variability and poor generalization across different imaging conditions. Although recent Deep Learning (DL) approaches improve detection performance, they often fail to effectively capture multi-scale pathological patterns and lack adaptive optimization of diagnostic decisions. These limitations highlight the need for an intelligent and automated diagnostic framework. This paper proposes an intelligent dental cyst prediction framework based on a Pyramid Swin Vision Transformer with Reinforcement Learning (PSViT-RL) based diagnostic optimization. The PSViT architecture hierarchically extracts multi-scale spatial features from dental radiographs, enabling accurate modeling of both lesion boundaries and global contextual information. To further enhance diagnostic performance, a RL module is incorporated to dynamically optimize diagnostic policies by adaptively adjusting feature importance and classification thresholds based on prediction confidence and clinical consistency. This hybrid learning strategy allows the system to progressively refine its decision-making beyond existing supervised learning. Experimental results demonstrate that the proposed PSViT-RL framework significantly outperforms existing CNN and transformer-based methods in terms of accuracy, sensitivity, and diagnostic consistency. The stable convergence and optimized decision-making achieved through RL indicate the framework's strong potential for deployment as a reliable clinical decision-support system in real-world dental imaging applications.

## 1. Introduction

Oral health has a significant impact on overall health, well-being, and medical expenses. In many countries, oral health remains a serious public health challenge that disrupts daily life and may lead to suffering, pain, or even loss of life. The 2019 Global Burden of Disease Study reported that dental diseases affect approximately 3.5 billion people worldwide. The prevalence of dental diseases is increasing rapidly, particularly in low- and middle-income countries, due to population growth and aging [1]. According to global reports, nearly 520 million children and about 2 billion adults suffer from cavities in their primary and permanent teeth. Conditions such as periodontitis, enamel caries, dental plaque, oral cancer, and periapical diseases are commonly diagnosed using dental X-rays, as they are often undetectable through routine oral examinations [2]. In this work, only four diseases, periodontal disease, periapical disease, pericoronal disease, and enamel disease, are considered. Advancements in medical imaging have made it a crucial component of modern healthcare by enabling accurate and non-invasive disease diagnosis and treatment planning. Although medical imaging has evolved rapidly, dental radiography has received comparatively less research attention due to the variability in anatomical structures across datasets [3]. Digital dental radiography is increasingly used by dental professionals to improve the detection, diagnosis, treatment, and monitoring of oral health conditions. Access to pre-diagnostic analytical support enables dentists to deliver more precise and effective patient care, supporting informed and timely clinical decision-making [4].

Image processing is an emerging technology used to extract meaningful information from images. A digital image is a two-dimensional representation composed of fixed values known as pixels (image elements). Pixel values encode attributes such as gray levels, intensity, color, and spatial information. Digital image processing refers to the manipulation of digital images using computational techniques. Low-level image processing includes basic operations such as noise removal, contrast enhancement, and image sharpening to improve image quality [5]. Mid-level processing involves tasks such as image segmentation, feature description, object recognition, and image classification. High-level processing focuses on extracting semantic information from recognized objects and is closely associated with computer vision. Dentistry is one of the healthcare domains that has received comparatively limited attention in terms of computational diagnosis [6]. Dental radiography remains the most widely used diagnostic modality, although radiographic images vary based on exposure levels, imaging location, and anatomical differences. These images assist dental practitioners in identifying structural abnormalities and provide critical information about tooth and bone formation, soft tissues, tooth loss, decay, Root Canal Therapy (RCT), cysts, and cavities that are not visible during routine clinical examinations [7].

Early diagnosis of dental diseases requires accurate detection and evaluation

of lesions caused by mineral loss in enamel and dentin. Individualized diagnosis demands assessment beyond visual signs and symptoms, including lesion activity and progression. This necessitates evaluating etiological and modifying factors through caries risk assessment to predict future disease development remains challenging. DL has emerged as a powerful tool in computer vision and healthcare applications [8]. In dentistry, DL plays a vital role in determining prosthetic rehabilitation levels and enabling automated CAD-based prosthetic design, significantly reducing fabrication time through digital workflows such as intraoral scanning and additive or subtractive manufacturing [9].

Partial dental crowns and indirect restorations cover only a portion of the tooth structure, represent some of the most complex tasks to automate in dental prosthetics. Inlays are typically designed on a case-by-case basis, whereas onlays may extend over one or more cusps. Although clinically feasible, onlay fabrication is more complex and time-consuming than inlay production and may negatively influence the long-term survival of in vivo prostheses [10]. Accurate classification of inlays and onlays is a critical initial step in the prosthetic fabrication process. Achieving high accuracy in 3D data classification remains challenging due to the geometric complexity of dental structures, variations introduced during rapid prototyping, and limitations of 3D data representations such as voxelization often fail to preserve surface smoothness and geometric properties. Several studies have explored automated techniques for dental caries detection [11]. Computer-aided approaches for evaluating carious lesions on permanent teeth have employed intensity- and texture-based feature extraction through region-of-interest identification followed by lesion labeling using the International Caries Detection and Assessment System (ICDAS). Classification was performed using correlation-based techniques with expert validation to improve accuracy. Proposed a neural network–based cavity recognition method integrated with feature mining and the Adaptive Dragonfly Algorithm was proposed, demonstrating the effectiveness of image processing and neural networks in accurately detecting dental cavities [12].

A computer-aided automated approach has been introduced to assess carious lesions on the surface of permanent teeth. The method employs a two-stage feature extraction process involving (a) identification of regions of interest and (b) classification of the detected regions using the International Caries Detection and Assessment System (ICDAS). Following feature extraction, a correlation-based technique is applied for lesion classification. To enhance performance, two dental experts validated the extracted occlusal surfaces [13]. Proposed a cavity recognition framework integrating feature mining, the Adaptive Dragonfly Algorithm, and a neural network–based classifier, achieving accurate detection of dental cavities. These studies demonstrate the effectiveness of image processing and neural network techniques in identifying and evaluating dental caries. The application of Artificial Intelligence (AI) for classifying pediatric dental diseases

using panoramic radiographs is a rapidly advancing research area with significant potential to improve diagnostic accuracy, reduce human error, and enhance patient outcomes [14]. However, challenges related to data availability, annotation consistency, and cross-population model validation must be addressed to realize its full potential. Although Convolutional Neural Network (CNN)–based image classification approaches have shown promising results, they often require extensive preprocessing and large annotated datasets. Advanced image-based deep learning models offer opportunities to improve accuracy, interpretability, and scalability in pediatric dental diagnosis [15]. Key contributions of the paper are as follows:

- Proposes a novel PSViT for multi-scale dental cyst prediction.
- Integrates RL to optimize diagnostic decisions dynamically and adaptively.
- Enhances robustness against radiographic variations and imaging noise across datasets.
- Reduces false positives and false negatives through reward-driven decision refinement.
- Demonstrates superior predictive performance compared to existing DL frameworks.

## 2. Related works

The images are pre-processed using histogram-based contrast enhancement on grayscale images. Tooth surfaces are segmented using region-growing techniques and the Circular Hough Transform (CHT). Morphological operations are applied to eliminate abnormal regions along the tooth boundaries. Image reconstruction is performed using the Haar wavelet transform. A total of 77 region-based and pixel-based features are extracted from both the upper and lower jaws. Feature ranking is conducted using the information gain ratio to select the most relevant features [16]. Five machine learning–based classifiers are employed for classification, achieving an accuracy of 86.3% and a sensitivity of 98.3%. The system is limited by a small number of images per category, lack of histological confirmation, and reliance on images of extracted teeth. Proposed a four-stage segmentation approach for dental caries using 40 periapical X-ray images, demonstrating that watershed segmentation outperforms min–max Euler-based segmentation [17].

Several studies have explored the use of 3D CNNs for classification tasks in medical imaging, including brain lesion segmentation, lung cancer detection, and dental treatment progression monitoring using 2D radiographs. These networks typically employ Conv3D layers, max pooling, and batch normalization, followed by fully connected layers for classification. In addition, some studies have integrated 3D CNN autoencoders into the processing pipeline. More recent research has combined both 2D and 3D CNN architectures with transformer-based models for brain image classification [18]. Comparative analyses of CNN and Vision Transformer (ViT) performance on 3D data and hybrid CNN–ViT approaches for 2D medical data classification, including gait analysis have demonstrated the advantages

of combining both architectures. Such hybrid strategies have not yet been extensively explored in digital dentistry. To address this gap, the present study proposes a transformer-based deep neural architecture integrating CNN and ViT models to classify 3D scans of prepared teeth for partial dental crowns [19]. The role of Explainable AI (XAI) techniques is investigated to improve interpretability and trust in model decisions. This study is among the first to apply ViTs and XAI for partial dental crown classification using 3D data and voxel-based representations during pre-processing. Comparative evaluations indicate that the hybrid CNN–ViT architecture achieves superior performance [20].

Proposed a novel image denoising algorithm based on transform-domain augmented sparse representation using collaborative filtering of three-dimensional groups. The resulting image is a three-dimensional approximation composed of jointly filtered image blocks. The method was evaluated on 50 panoramic X-ray images, with accuracy validated by medical specialists. However, key limitations include preserving the overall image appearance, maintaining diagnostically relevant regions, and detecting low-contrast features [21]. The approach is inspired by Blind Source Separation (BSS), modeling panoramic images as a combination of background information, diagnostic content, and noise. The method achieved a Peak Signal-to-Noise Ratio (PSNR) of 35 dB. Designed and evaluated an X-ray image denoising system using Generative Adversarial Networks (GANs) [22]. The DL based approach categorizes teeth into molars, premolars, canines, and incisors using cone-beam CT images through an AlexNet-based architecture. Although CNNs are effective for multi-class identification and characterization, they typically require large-scale training datasets to achieve high performance when applied directly [23].

Introduced a variant of the Double-Scale Nonlinear Thresholding (DNT) method for fine image segmentation based on grayscale intensity values within vessel-support regions. A major limitation of this method lies in selecting appropriate threshold values, particularly when processing noisy images, which can adversely affect segmentation accuracy. Developed a mixed-scale dense convolutional neural network for bone segmentation in metal artefact–affected Cone-Beam Computed Tomography (CBCT) images [24]. A dataset of 20 dental CBCT scans was used for training, where bone structures were initially segmented using global thresholding by an expert medical engineer, followed by manual removal of residual noise and metal artefacts. Although notable progress has been achieved in deep learning–based dental image analysis, existing dental cyst prediction systems still suffer from several critical limitations [25]. Most existing approaches rely predominantly on conventional CNN-based architectures are insufficient in capturing multi-scale features and long-range contextual dependencies required to differentiate visually similar cystic and non-cystic regions. Prior studies primarily focus on detection or classification tasks, while neglecting predictive risk assessment and

adaptive decision-making essential for clinical reliability [26]. Existing models adopt static learning paradigms and lack reinforcement or feedback-driven optimization to refine diagnostic policies during inference. Challenges such as limited explainability, poor generalization across heterogeneous imaging modalities (e.g., OPG and CBCT), and variability in imaging conditions remain unresolved [27]. The absence of clinically informed constraints and dynamic confidence calibration further contributes to increased false diagnosis rates. These limitations highlight the need for an intelligent, adaptive, and context-aware dental cyst prediction framework that integrates hierarchical vision transformers with RL based diagnostic optimization [28].

## 3. Materials and methods

The proposed intelligent dental cyst prediction framework integrates PSViT-RL based diagnostic optimization to analyse dental radiographic images such as OPG and CBCT scans shown in Figure 1. Initially, the input images undergo pre-processing to enhance contrast and suppress noise, enabling clearer visualization of cystic regions. The enhanced images are partitioned into non-overlapping patches and fed into the PSViT performs hierarchical feature learning across multiple stages. This enables effective extraction of fine-grained local lesion features and broader anatomical context, resulting in robust multi-scale feature representations for accurate cyst identification. The aggregated multi-scale features are forwarded to a primary classification module that predicts the presence or absence of dental cysts along with associated confidence scores. To further improve diagnostic reliability, an RL module is incorporated to optimize decision-making dynamically. The RL agent evaluates prediction outcomes using reward signals based on accuracy, sensitivity, and clinical consistency, and adaptively adjusts decision thresholds and feature weighting in real time. This feedback-driven optimization reduces false positives and false negatives while improving generalization across imaging modalities. Overall, the synergistic integration of PSViT-RL establishes a reliable and intelligent clinical decision-support system for automated dental cyst detection and risk assessment.

### 3.1 Dataset Description

The dental cyst dataset used in this research comprises Orthopantomogram (OPG) and Cone Beam Computed Tomography (CBCT) radiographic images collected from clinical dental centers and publicly accessible repositories shown in Table 1. The dataset includes images of confirmed cyst cases as well as normal (non-cyst) cases, enabling the model to learn both pathological and normal patterns effectively. All images were down sampled to a uniform resolution and converted to grayscale to ensure consistency across samples.

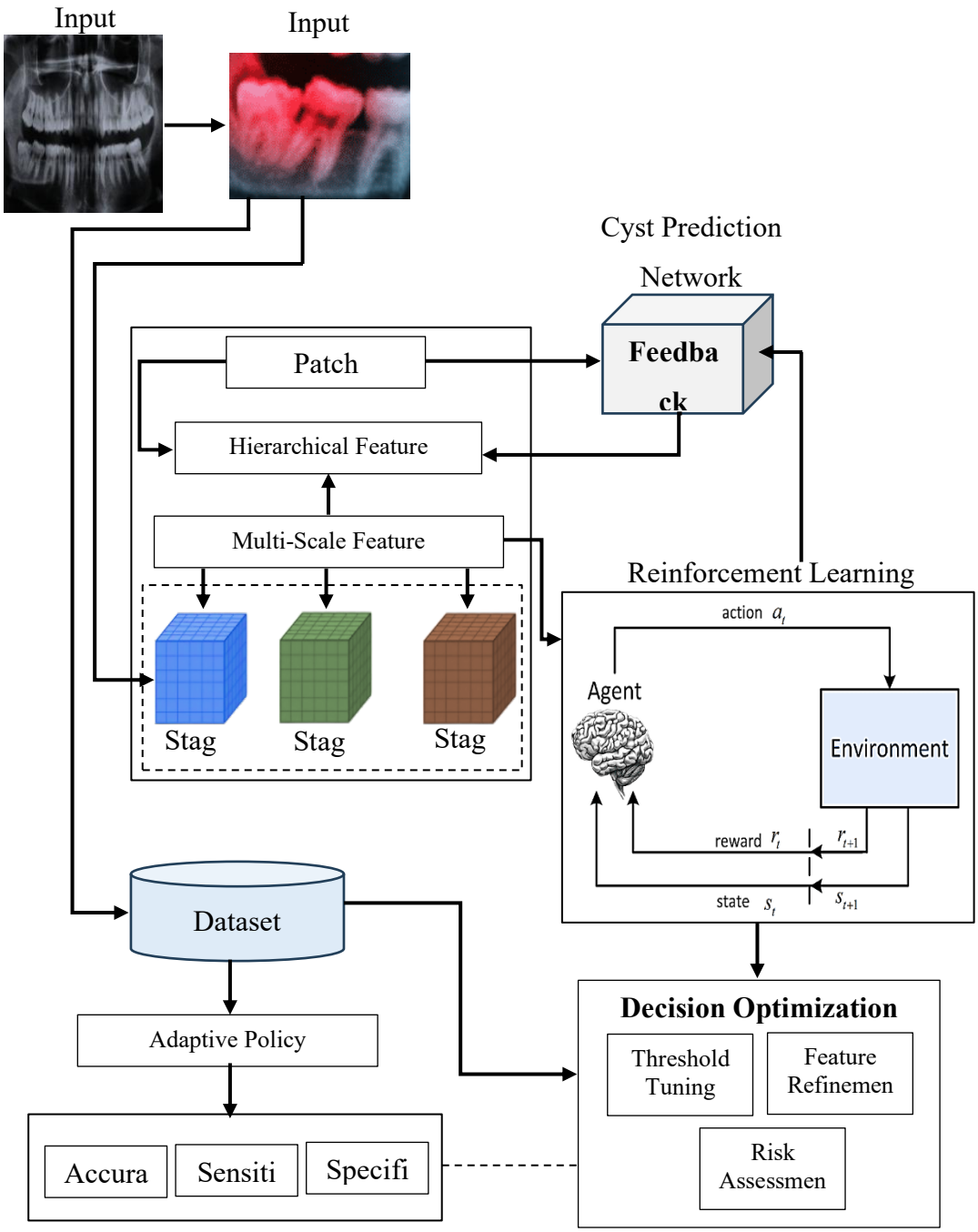


**Fig.1 Proposed Architecture**

The dataset was annotated by experienced oral radiologists identified cyst regions and assigned class labels. These annotations were further reviewed to ensure high-quality ground truth. Prior to model training, pre-processing steps such as noise removal, contrast enhancement, image

normalization, and image resizing were applied to improve image quality and enhance learning performance.

**Table 1: Dataset Description**

| Dataset Attribute | Description |
|---|---|
| Imaging Modality | OPG and CBCT images |
| Dataset Source | Collected from partnered dental hospitals and publicly available dental radiography repositories |
| Total Images | 1,200 dental radiographic images |
| Cyst Cases | 650 images containing confirmed dental cysts |
| Non-Cyst Cases | 550 normal or non-cyst pathological images |
| Image Resolution | Standardized to 224 × 224 pixels |
| Image Format | Grayscale PNG/JPEG |
| Annotation Type | Expert-annotated bounding regions and class labels (cyst / non-cyst) |
| Ground Truth | Verified by experienced oral radiologists |
| Data Split | 70% Training, 15% Validation, 15% Testing |
| Preprocessing Steps | Noise reduction, contrast enhancement, normalization, resizing |
| Augmentation Techniques | Rotation, flipping, scaling, intensity variation |
| Class Balance Strategy | Augmentation and weighted loss to handle class imbalance |

**Table 2: Sample Data**

| Sample ID | Imaging Modality | Class Label | Cyst Type | Affected Region | Annotation | Image Size |
|---|---|---|---|---|---|---|
| IMG_001 | OPG | Cyst | Radicular Cyst | Mandibular Molar Region | Expert Labeled | 224 × 224 |
| IMG_002 | CBCT | Non-Cyst | – | Normal Jaw Structure | Expert Verified | 224 × 224 |
| IMG_003 | OPG | Cyst | Dentigerous Cyst | Impacted Third Molar | Expert Labeled | 224 × 224 |
| IMG_004 | CBCT | Cyst | Odontogenic Cyst | Maxillary Premolar Area | Expert Labeled | 224 × 224 |

| | | | | | | |
|---|---|---|---|---|---|---|
| IMG_005 | OPG | Non-Cyst | – | Healthy Dental Anatomy | Expert Verified | 224 × 224 |

The sample dental radiographic images, including both cyst-positive and non-cyst cases, represent the diversity of the dataset shown in Table 2. Each sample corresponds to an individual patient scan and is labeled based on professional clinical diagnosis. Cyst-positive images typically exhibit distinct radiolucent regions near tooth roots or affected teeth, whereas non-cyst images show normal anatomy without pathological abnormalities. Metadata such as image ID, imaging modality (OPG or CBCT), class label, and annotation status are recorded for each sample. These samples demonstrate variations in cyst size, shape, and location, highlighting the need for multi-scale feature extraction and validating the robustness of the proposed framework.

The 3D CBCT images clearly differentiate jaws with and without dental cysts across axial, coronal, and sagittal views shown in Figure 2. In cyst-positive cases, the radiographs show a distinct radiolucent cavity within the jawbone, indicating abnormal pathological growth and disruption of the normal bone structure. In contrast, cyst-negative cases exhibit uniform bone density, an intact trabecular pattern, and proper tooth alignment. These multi-view 3D representations provide essential volumetric and contextual information, which supports accurate detection and classification of dental cysts.

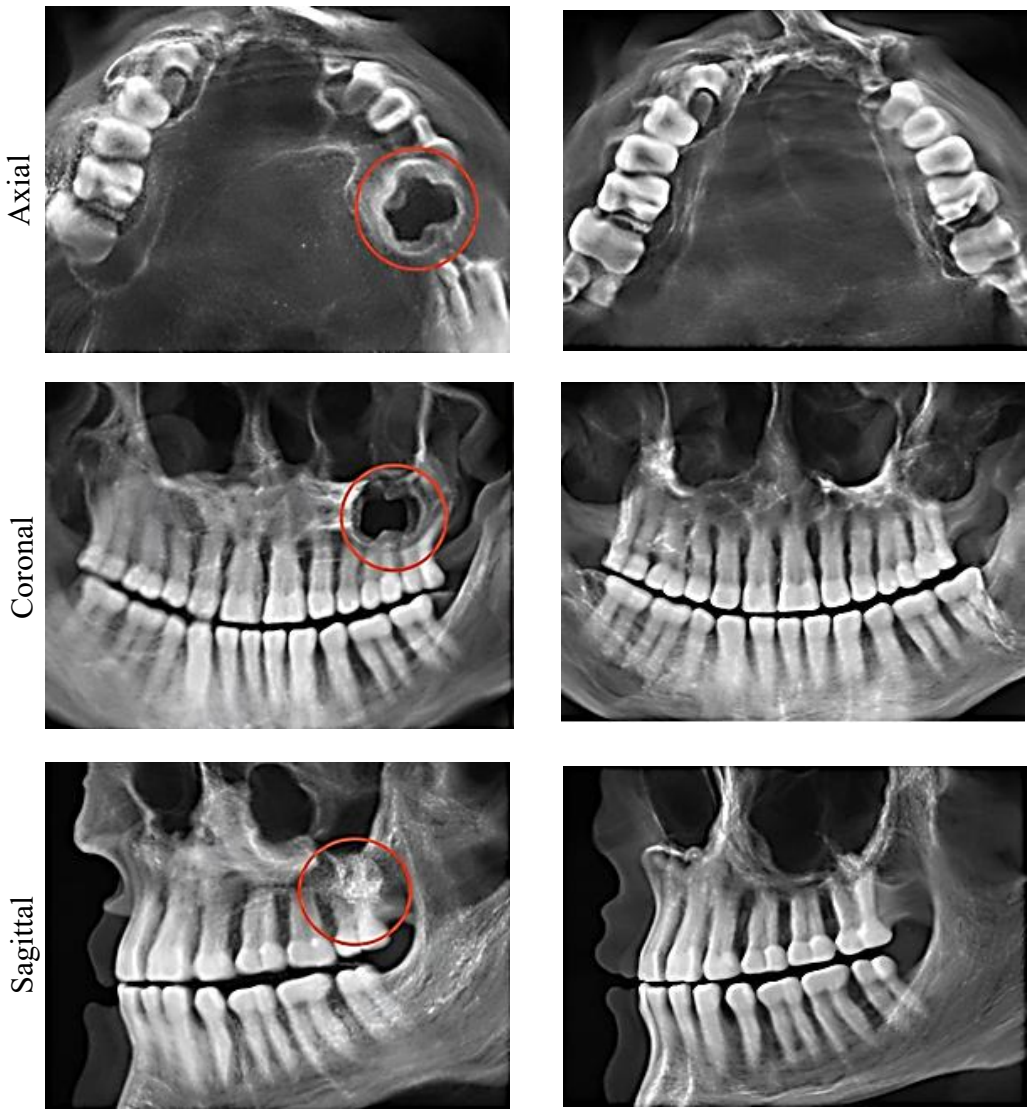


**Fig.2 3D CBCT Sample Image**

### 3.2 Pre-processing

Pre-processing of images will be an essential component of the proposed framework of intelligent dental cyst prediction because the dental radiographs can be characterized by poor contrast, noise, and intensities. With adequate pre-processing, the PSViT can have stable learning and greater diagnostically relevant structures. First, image resizing and grayscale normalization are applied to standardize all input images. Given an input image $X \in R^{H \times W}$ it is resized to a fixed resolution $224 \times 224$ using bilinear interpolation. Pixel intensities are normalized to the range [0, 1] as

$$X_{norm}(i,j) = \frac{X(i,j)}{255} \quad (1)$$

This normalization stabilizes gradient updates and accelerates convergence during training. Second, noise reduction is performed to suppress speckle and acquisition noise common in OPG and CBCT images. A Gaussian filter is applied as

$$X_{den}(i,j) = \sum_{x=-k}^{k} \sum_{y=-k}^{k} X_{norm}(i-x, j-y) G(x,y) \quad (2)$$

Where,

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

This step preserves anatomical boundaries while removing high-frequency noise that may mislead feature extraction. Third, contrast enhancement is applied to highlight cystic regions, which typically appear as radiolucent areas. Histogram equalization or Contrast Limited Adaptive Histogram Equalization (CLAHE) redistributes intensity values using the Cumulative Distribution Function (CDF):

$$X_{enh}(i,j) = (L-1) \sum_{n=0}^{X_{den}(i,j)} p(n) \quad (4)$$

Where L is the number of gray levels and p(n) is the normalized histogram. This improves visibility of subtle cyst boundaries.

Fourth, intensity standardization (Z-score normalization) is applied to reduce inter-image variability:

$$X_{std}(i,j) = \frac{X_{enh}(i,j) - \mu}{\sigma} \quad (5)$$

Where $\mu$ and $\sigma$ are the mean and standard deviation of pixel intensities. This ensures consistent intensity distribution across patients and imaging devices. Finally, data augmentation is used to improve generalization. Augmented images

$X'$ are generated through transformations such as rotation and flipping:

$$X' = T(X_{std}),\ T \in \{rotate(\theta), flip, scale\} \quad (6)$$

These pre-processing steps collectively enhance image quality, emphasize cystic patterns, and provide robust, standardized inputs for effective multi-scale feature learning in the PSViT.

### 3.3 Data Augmentation for Dental Cyst Prediction

It is an important pre-processing step that artificially increases the size and diversity of the training dataset, thereby enhancing the robustness and generalization capability of deep learning models such as PSViT. In dental imaging, variations in patient anatomy, imaging angles, and noise can significantly affect model performance. To address these challenges, data augmentation techniques are applied to generate new training samples while preserving the anatomical integrity of teeth and cysts. Common augmentation operations include rotation, flipping, scaling, translation, contrast adjustment, and Gaussian noise. Let the original image be represented as $X \in R^{H \times W \times C}$ where H, W, and C denote height, width, and channels. A transformation T from the augmentation set $\mathcal{T}$ generates an augmented image $X'$:

$$X' = T(X),\ T \in \mathcal{T} \quad (7)$$

For example, a rotation by angle $\theta$ can be represented using an affine transformation matrix

$$X' = R_\theta . X \quad (8)$$

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (9)$$

Similarly, adding Gaussian noise $N(0, \sigma^2)$ to the image pixels enhances model robustness to imaging artifacts:

$$X' = X + N(0, \sigma^2)$$

(10)

Data augmentation effectively increases the number of training examples N to N' without the need for additional data collection:

$$N' = N \times |T|$$

(11)

By exposing the PSViT to these varied augmented images, the model learns scale-invariant, rotation-invariant, and noise-robust representations is essential for accurately detecting dental cysts across different patients and imaging conditions. Figure 3 illustrates the data augmentation process applied to CBCT dental images. The first column shows the original image with cyst regions marked, while the second column presents augmented versions using rotation, flipping, zooming, and contrast and brightness adjustments. The third column displays the corresponding transformed ground truth masks, ensuring label consistency. Additional elastic deformations simulate anatomical variations. This process increases dataset diversity, reduces overfitting, and helps the PSViT learn robust and generalizable features for accurate cyst detection.

### 3.4 Pyramid Swin Vision Transformer architecture for feature extraction

The PSViT framework begins by reconstructing a 3D dental CBCT volume into 2D slices, which are divided into fixed-size, non-overlapping patches shown in Figure 4. These patches are flattened and projected into a latent feature space via linear projection, forming a sequence of embedded tokens suitable for transformer processing. The tokens undergo Window-based Multi-Head Self-Attention (W-MSA), where attention scores are computed from the similarity between query and key vectors, normalized with softmax, to capture local contextual relationships. Shifted Window MSA (SW-MSA) is applied between alternating windows to enable information exchange across neighboring windows, improving global feature representation. Multi-scale hierarchical feature maps are generated through patch merging layers, which concatenate and linearly transform neighboring patch features, reducing spatial resolution while increasing channel depth. Shallow layers capture fine cyst boundaries

and textures, while deeper layers encode high-level semantic features, including cyst shape, size, and anatomical context. The framework consists of four main stages: patch partitioning & embedding, hierarchical attention processing, patch merging, and multi-scale feature map generation.

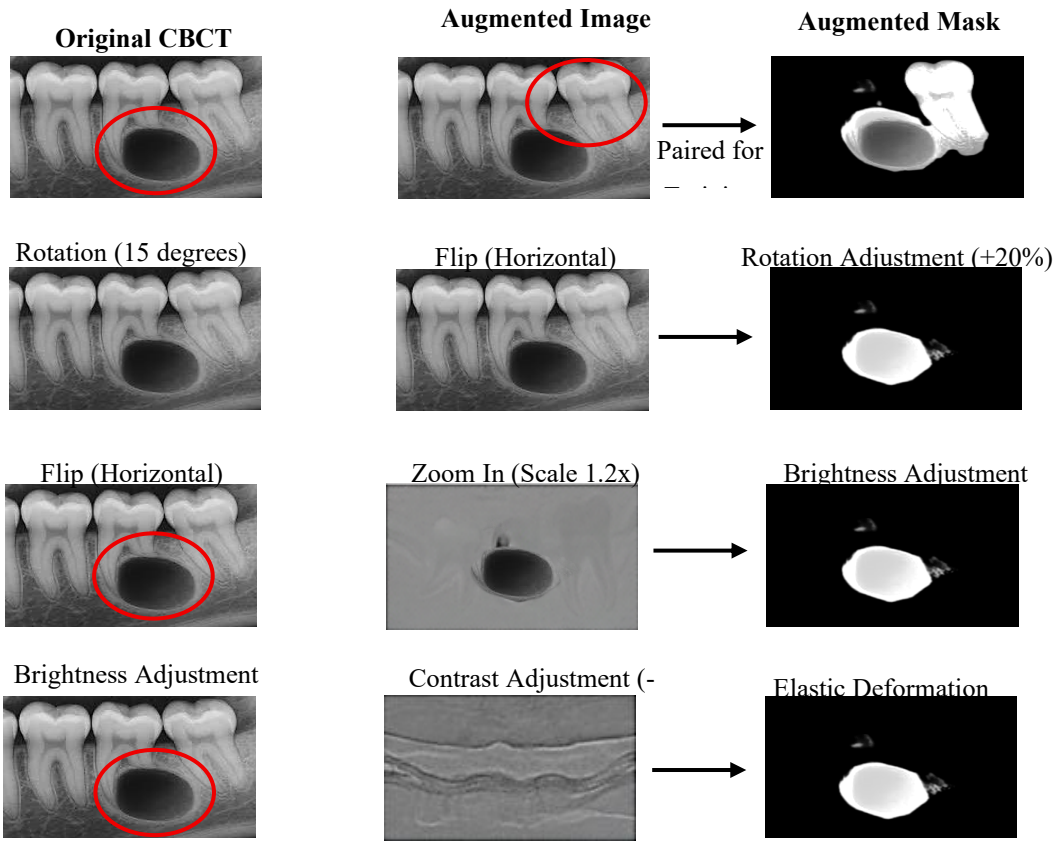


**Fig. 3 Data Augmentation Image**

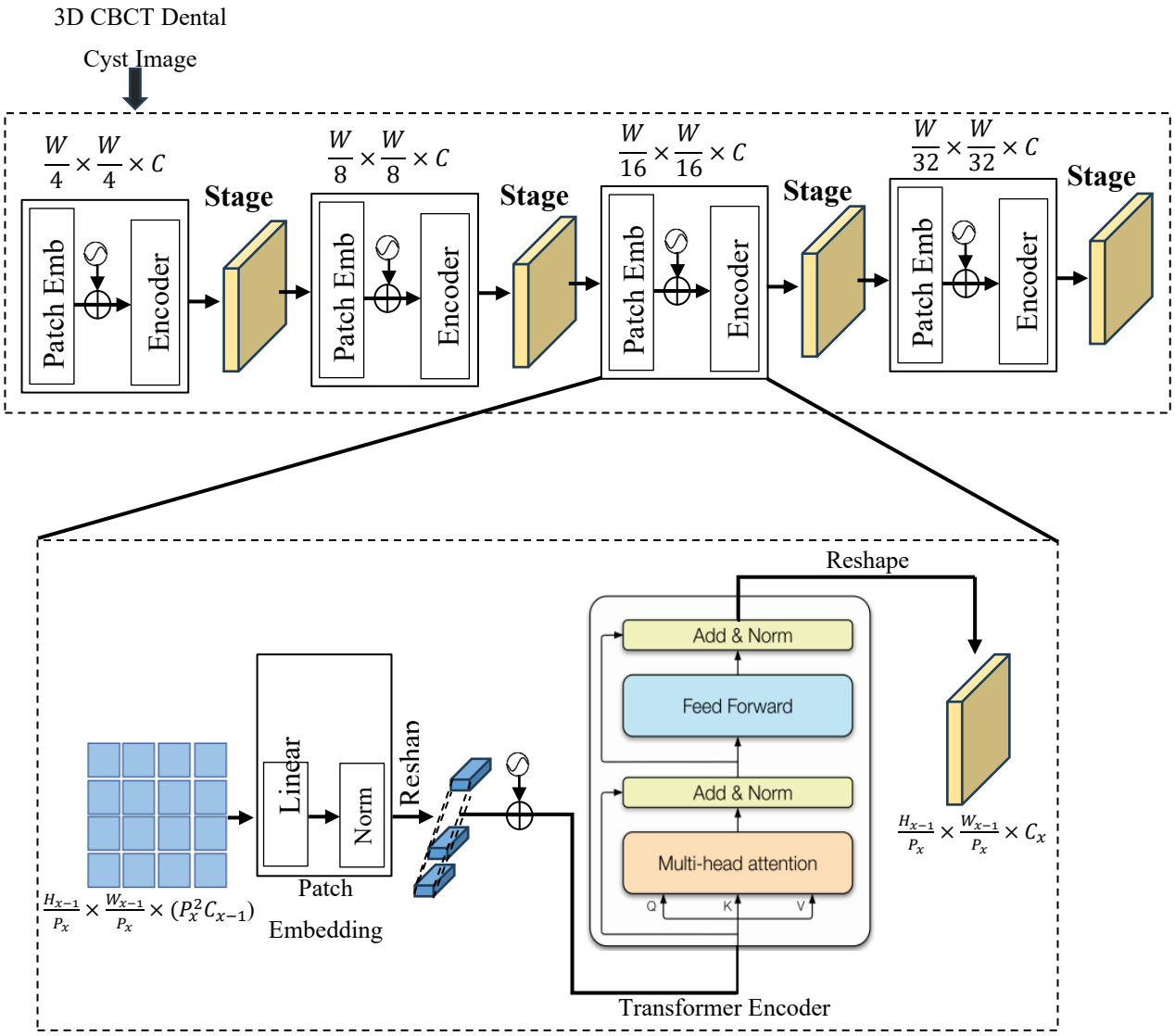


**Figure 4:** Representation of PSViT architecture for feature extraction

**Stage 1: Patch Partitioning and Embedding**

The input to the model is a 3D CBCT dental cyst image typically has high resolution and volumetric information. To reduce computational complexity, the image is first divided into smaller non-overlapping patches of size $\frac{W}{4} \times \frac{W}{4}$ assuming spatial dimensions H x W). If the original image has dimensions HW then the number of patches N is computed as:

$$N = \frac{H.W}{P^2} \quad (12)$$

Where P = 4 is the patch size. Each patch is CBCT 3D vector $i_p \in R^{P^2.C}$ where C is the number of channels in the input image. The patch embedding objects each flattened patch to a higher- dimensional feature space using a linear transformation:

$$z_0 = i_p W_e + b_e \quad (13)$$

Here, $w_e \in R^{(P^2.C)\times D}$ is the learnable embedding weight matrix, $b_e \in R^D$ is the bias, and D is the embedding dimension.

**Stage 2 - Local and Global Attention**

The embedded patches are passed through Stage 2 applies W-MSA to capture local dependencies. The attention mechanism operates within non-overlapping windows, reducing computational complexity from $O(N^2)\ to\ O(M^2.\frac{N}{M^2}) = O(N.M^2)$ where M is the window size. The W-MSA output is calculated as:

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (14)$$

Where $Q = XW_Q,\ K = IW_K,\ V = IW_V$ are the query, key, and value matrices, and $d_k$ is the dimension of each head. Stage 2 reduces spatial dimensions to $\frac{W}{4}\times\frac{W}{4}$ and increases feature channels to C.

**Stage 3: Patch Merging**

After W-MSA, Patch Merging reduces spatial resolution and doubles the channel dimension. For a given stage s, the merged feature map is:

$$I_{s+1} = Concat(i_{2x,2y}, i_{2x+1,2y}, i_{2x,2y+1}, i_{2x+1,2y+1}) \quad (15)$$

Here, $W_m \in R^{4C\times 2C}$ is a linear projection, which aggregates $2\times 2$ neighboring patches, reducing the resolution while enriching channel-wise representation. Stage 3 applies SW-MSA shifts windows across patches to capture cross-window interactions. The resolution becomes $\frac{W}{16}\times\frac{W}{16}$ and channels increase to 4C.

**Stage 4: Multi-Scale Feature Maps**

It continues SW-MSA and W-MSA operations, followed by patch merging. The final stage reduces the spatial dimension to $\frac{W}{32}\times\frac{W}{32}$ while increasing feature channels to 8C. This hierarchical representation preserves fine-grained local structures while capturing global context, producing multi-scale feature maps suitable for downstream tasks like detection,

segmentation, or classification of dental cysts.

$$F_{multi-scale} = \{X_2, X_3, X_4\} \quad (16)$$

Where $X_2, X_3, X_4$ correspond to feature maps from Stage 2, Stage 3, and Stage 4, respectively.

In essence, this framework efficiently combines:

- Patch partitioning to reduce input complexity.
- Window-based self-attention for local and shifted window attention for global interactions.
- Hierarchical patch merging to extract multi-scale representations.
- Output multi-scale feature maps for effective 3D CBCT dental cyst analysis.

This design balances computational efficiency with representational power, making it highly suitable for medical image analysis where both local details and global structures are crucial.

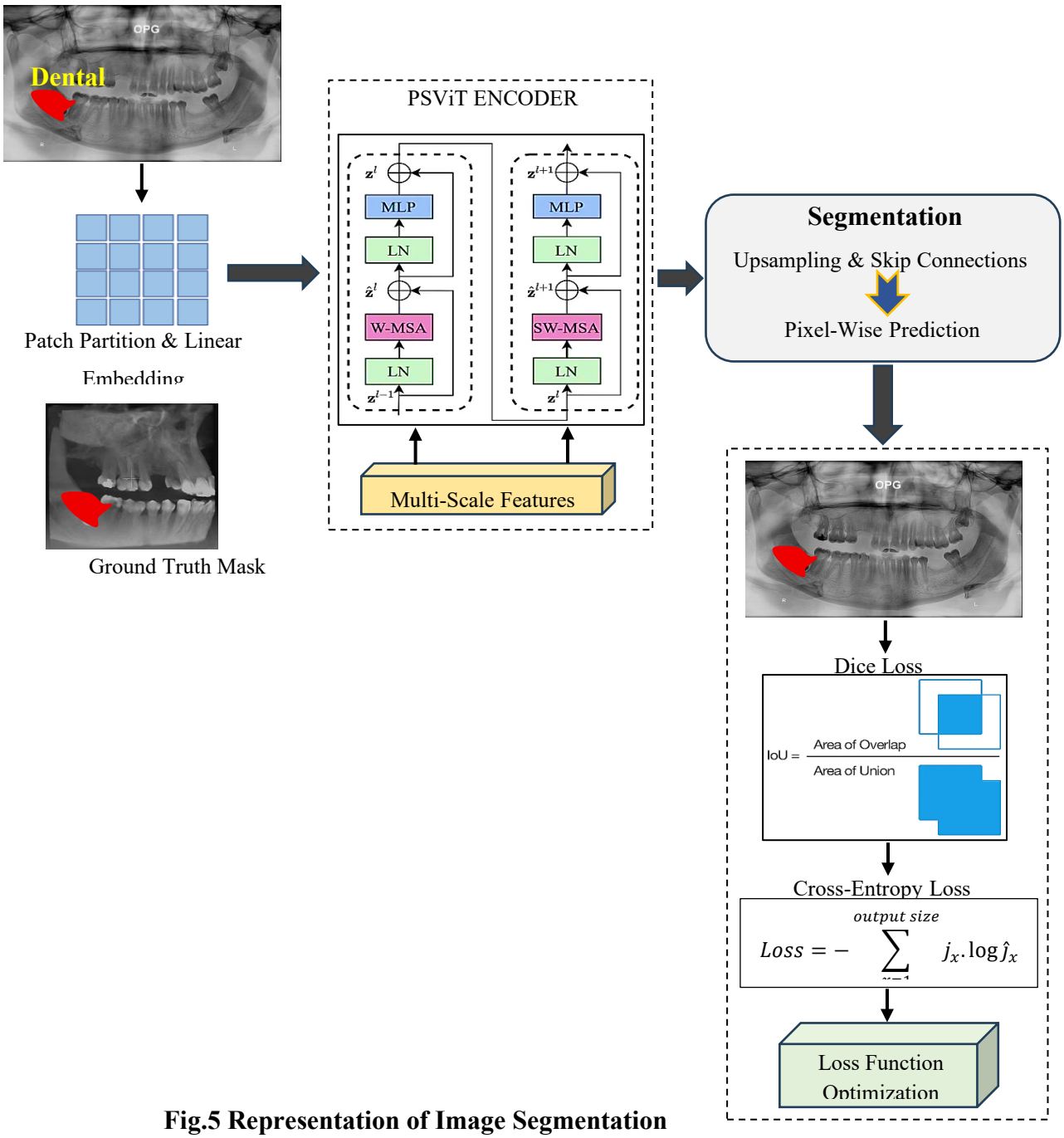


**Fig.5 Representation of Image Segmentation**

### 3.5 Image segmentation using PSViT with segmentation head

The proposed learning model for detecting oral cysts is based on a PSViT integrated with a segmentation head to produce fine-grained cyst segmentation from CBCT or dental radiographic images shown in Figure 5. PSViT exploits a pyramid architecture to extract hierarchical multi-scale spatial features by partitioning

images into patches and applying transformer-based self-attention, enabling effective modeling of both local details and global contextual information. The extracted features are subsequently passed to a segmentation head, typically composed of upsampling layers with skip connections, which generates pixel-wise predictions in the form of a binary cyst mask. During training, the overlap between predicted and ground-truth masks is optimized using loss functions such as Dice loss or cross-entropy loss. This hybrid design enables accurate delineation of cyst boundaries even within complex jaw structures by combining the global contextual awareness of transformers with precise spatial localization. The resulting framework provides an automated, interpretable, and clinically relevant tool for dental cyst detection, supporting diagnosis, treatment planning, and reducing reliance on manual annotations.

The PSViT architecture with a segmentation head is designed for precise dental cyst detection in CBCT or dental X-ray images. The workflow involves the following steps:

**Input and Patch Embedding:** The input dental image $X \in R^{H\times W\times C}$ is first divided into non-overlapping patches of size $P \times P$. Each patch is flattened and projected to a feature vector:

$$i_p = Flatten(X_p)W_e + b_e, \quad i_p \in R^D \tag{17}$$

Where $W_e \in R^{(P^2.C)\times D}$ is the embedding matrix, D is the embedding dimension, and $X_p$ is the $p^{th}$ patch.

**Pyramid Split Transformer Encoding:** PSViT uses a multi-scale pyramid split to capture both local and global context. At each pyramid level l, features are split into smaller patches and passed through transformer layers:

$$Z^{(l)} = MSA(LN(X^{(l-1)}) + I^{(l-1)} \tag{18}$$

$$X^{(l)} = MLP(LN(Z^{(l)}) + Z^{(l)} \tag{19}$$

Where, MSA = Multi-Head Self-Attention, capturing global dependencies, LN Layer Normalization, MLP Feed-Forward Network, Residual connections ensure stable gradient flow. At each level, multi-scale features are generated:

$$F_{multi-scale} = \bigcup_{l=1}^{L} I^{(l)} \tag{20}$$

**Segmentation Head:** The segmentation head converts the multi-scale features into a pixel-wise prediction map. This usually involves upsampling (deconvolution) layers with skip connections from lower

pyramid levels: $S = \sigma(Upsample(F_{multi-scale}) + F_{skip})$ (21)

Here, $\sigma$ is the sigmoid (or softmax) function for pixel-level probability, and $F_{skip}$ provides spatial details from earlier layers.

**Loss Function and Optimization:** The predicted segmentation mask $\hat{J}$ is compared to the ground truth mask $J$ using a combination of Dice Loss and Cross-Entropy Loss:

$$L_{Dice} = 1 - \frac{2\sum_x \hat{J}_x J_x}{\sum_x \hat{J}_x + \sum_x J_x + \epsilon} \quad (22)$$

$$L_{CE} = -\sum_x [J_x \log(\hat{J}_x) + (1 - J_x) \log(1 - \hat{J}_x)] \quad (23)$$

$$L_{total} = L_{Dice} + L_{CE} \quad (24)$$

The model is trained by minimizing $L_{total}$ using gradient descent:

$$\theta \leftarrow \theta - \alpha \nabla_\theta L_{total} \quad (25)$$

Where are all trainable parameters and a is the learning rate

The PSViT-based dental cyst segmentation system is a system combining PSViT and a segmentation head to perform pixel-wise cyst location. The received dental image is split into patches, which are embedded and encoded by using the encoder PSViT to extract multi-scale features, both local and global context by self-attention. The model is trained with the combination of both dice loss and cross-entropy loss and optimizes the pixel-level accuracy. In this architecture, dental cysts are automatically and interpretably segmented to provide a stronger diagnosis and treatment plans.

### 3.6 RL–driven diagnostic optimization

RL-based diagnostic optimization is based on the idea that the principles of RL are used to optimize clinical diagnostics shown in Figure 6. The diagnostic system is an intelligent agent, which interrelates with the clinical environment comprises patient data, lab results, and medical imaging. In every decision point, the agent selects an action, whether it is the recommendation of a test, an order of imaging, or an alternative line of treatment, and gets feedback on how the action is effective in being diagnostic with respect to cost and patient safety. The agent process improvises with time, acquiring the best sequences of diagnostic steps to be used, dynamically adjusting to patient profiles and medical conditions. The approach allows to find more effective,

customized, and precise diagnostic plans, minimize irrelevant processes and enhance patient outcomes, as well as maximize clinical workflow.

RL-Driven Diagnostic Optimization applies RL principles to improve clinical decision-making by modeling diagnostics as an interaction between an agent and the environment. The agent represents the diagnostic system, which observes the state of the patient $s_t$ including clinical data, lab results, and medical imaging, and chooses an action $a_t$ such as recommending a test, ordering imaging, or suggesting a treatment plan. The environment provides feedback in the form of a reward $r_t$ based on diagnostic accuracy, safety, and cost-effectiveness, and transitions to a new state $s_{t+1}$.

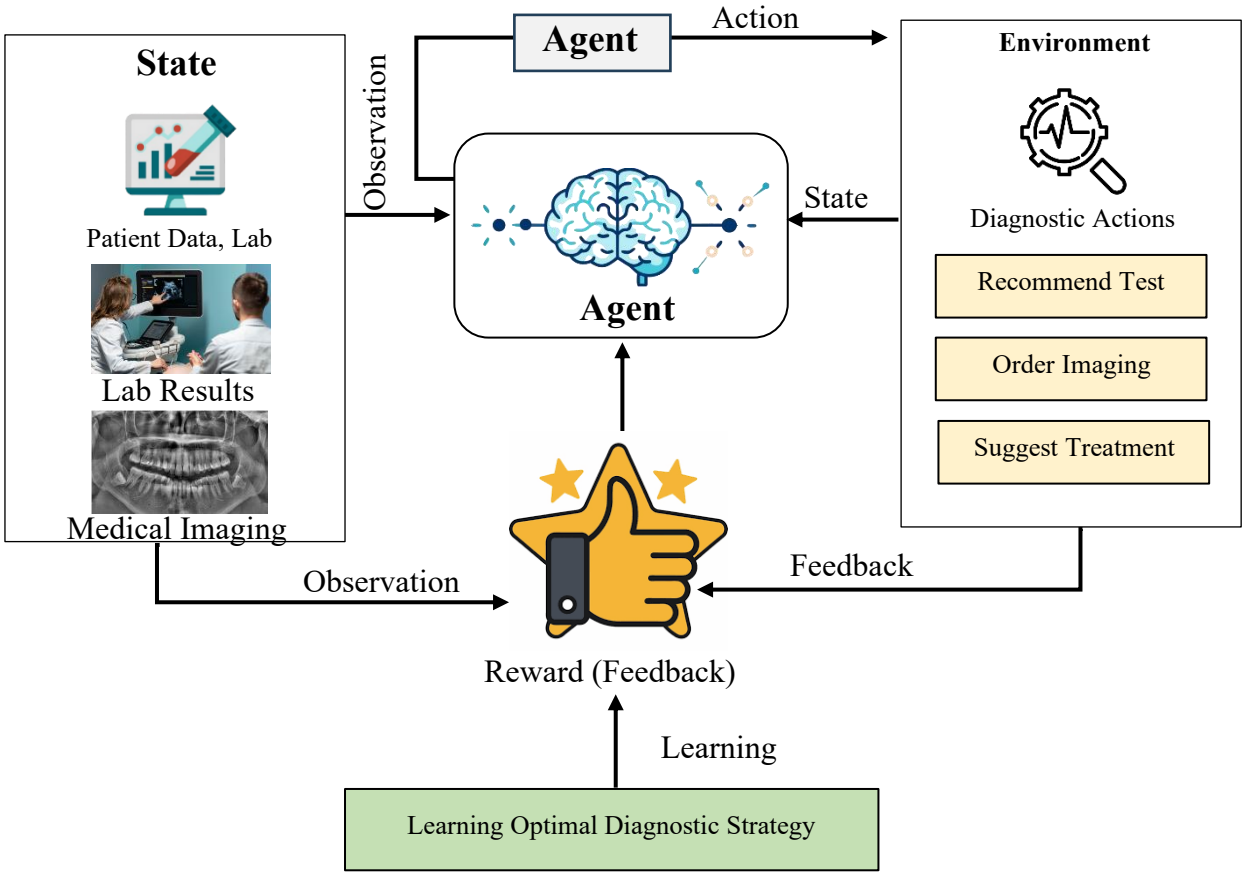


**Fig.6 Representation of Reinforcement Learning Driven Diagnostic Optimization**

The objective of the agent is to learn an optimal policy $\pi^*(s)$ that maximizes the expected cumulative reward over time. The agent estimates the action-value function $Q(s,a)$ representing the expected cumulative reward starting from states and taking action a. The Q-values are updated iteratively using the Bellman equation:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[ r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \right] \quad (26)$$

Where $\alpha$ is the learning rate controlling the update magnitude. $R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$ (27)

Where $\gamma$ is the discount factor determining the importance of future rewards.

By interacting with the environment and updating $Q(s, a)$ the agent learns the policy that maps patient states to optimal diagnostic actions: $\pi^*(s) = \arg\max_a Q(s, a)$ (28)

Through this process, the RL-driven system dynamically refines diagnostic strategies, reducing unnecessary tests while improving accuracy, efficiency, and patient outcomes. Integrating patient-specific observations with reward-driven learning allows adaptive, personalized, and optimized decision-making in complex clinical workflows.

**Algorithm: Intelligent Dental Cyst Prediction Framework**

Input: Dental CBCT/X-ray images X, patient metadata M

Output: Segmented dental cyst mask $\hat{J}$, optimized diagnostic actions

**Step 1: Image Pre-processing and Patch Embedding**

1. Divide input image $X \in R^{H \times W \times C}$ into patches of size $P \times P$

2. Flatten each patch and project to feature embedding $i_p$ *using Equation* (17)

**Step 2: Multi-Scale Feature Extraction using PSViT**

1. Pass the patch embedding's through pyramid-level transformer layers to extract hierarchical features.

2. Apply multi-head self-attention (MSA) and feed-forward networks (MLP) with residual connections using Equations (18) and (19)

3. Combine multi-scale features using Equation (20)

**Step 3: Segmentation Head for Dental Cyst Mask**

1. Upsample multi-scale features with skip connections to generate pixel-wise predictions using Equation (21)

2. Compute segmentation loss using Dice Loss and Cross-Entropy Loss using Equations (22) – (24)

**Step 4: RL-Driven Diagnostic Optimization**

1. Define state $s_t$ from patient metadata and segmentation results.

2. Agent selects diagnostic action $a_t$ (e.g., recommend additional tests or imaging).

3. Environment returns reward $r_t$ based on diagnostic accuracy, cost, and safety.

4. Update Q-values iteratively using the Bellman Equation (26)

5. Derive the optimal diagnostic policy using Equation (28)

**Step 5: Output**

1. Segmented dental cyst mask $\hat{f}$ for treatment planning.

2. Optimized diagnostic actions $a_t$ for patient-specific decision support.

The intelligent dental cyst prediction framework integrates the Pyramid Swin Vision Transformer (PSViT) with reinforcement learning (RL) to support optimized diagnostic decision-making while leveraging the high accuracy of PSViT-based multi-scale segmentation features. PSViT extracts hierarchical features from dental images and generates precise cyst masks through a dedicated segmentation head. The resulting segmentation outputs, along with patient data, are provided to the RL module, where diagnostic actions—such as test recommendations—are guided by reward signals based on accuracy, safety, and efficiency. This integrated approach enables automated, personalized, and effective diagnostics, resulting in improved cyst detection, reduced unnecessary procedures, and enhanced clinical outcomes.

## 4. Results and Discussions

The experimental setup of the intelligent dental cyst prediction framework comprises imaging, computational, and evaluation components. Clinical CBCT and dental X-ray images are collected and pre-processed through noise removal, normalization, and resizing to standard dimensions suitable for PSViT input. The PSViT model is trained using annotated cyst masks with Dice Loss and Cross-Entropy Loss. Simultaneously, segmentation outputs and patient metadata are fed into a reinforcement learning (RL) module, where the agent explores diagnostic actions and receives reward feedback based on accuracy, cost, and safety. Experiments are conducted on high-performance GPUs, and key hyperparameters such as patch size, learning rate, batch size, and RL parameters (discount factor, learning rate, and exploration rate) are optimized. Model performance is evaluated using Dice coefficient, IoU, accuracy, sensitivity, specificity, and RL policy efficiency through cumulative reward maximization.

**Table 3. Hyper-Parameter Settings**

| Module | Hyperparameter | Value / Setting |
|---|---|---|
| PSViT Encoder | Patch size (P) | 4 × 4 |
| | Embedding dimension (D) | 96 |
| | Number of pyramid levels (L) | 4 |
| | Number of attention heads | 8 |
| | Learning rate (α) | $1e^{-4}$ |
| | Batch size | 16 |
| | Number of epochs | 100 |
| Segmentation Head | Upsampling method | Bilinear + skip connections |
| | Activation function | Sigmoid / Softmax |
| RL | Discount factor (γ) | 0.9 |
| | Learning rate (α) | 0.01 |
| | Exploration rate (ε) | 0.1 → 0.01 (decay) |
| | Number of episodes | 500 |
| Loss Functions | Dice Loss weight | 0.5 |
| | Cross-Entropy Loss weight | 0.5 |

Hyper parameters of the framework are optimally configured. The PSViT encoder models have small patches, multi-level attention, and 96-dimensional embedding, and are optimized on a batch of 16 images with a single hundred epochs and a learning rate of $1e^{-4}$ shown in Table 3. Upsampling and skip connections as well as sigmoid/softmax activation are used by the segmentation head. It applies the RL module with a discount factor of 0.9, learning rate 0.01 and a reducing exploration rate as the number of episodes increases (500). Dice and Cross-Entropy losses are used with the same weight to result in proper cyst segmentation.

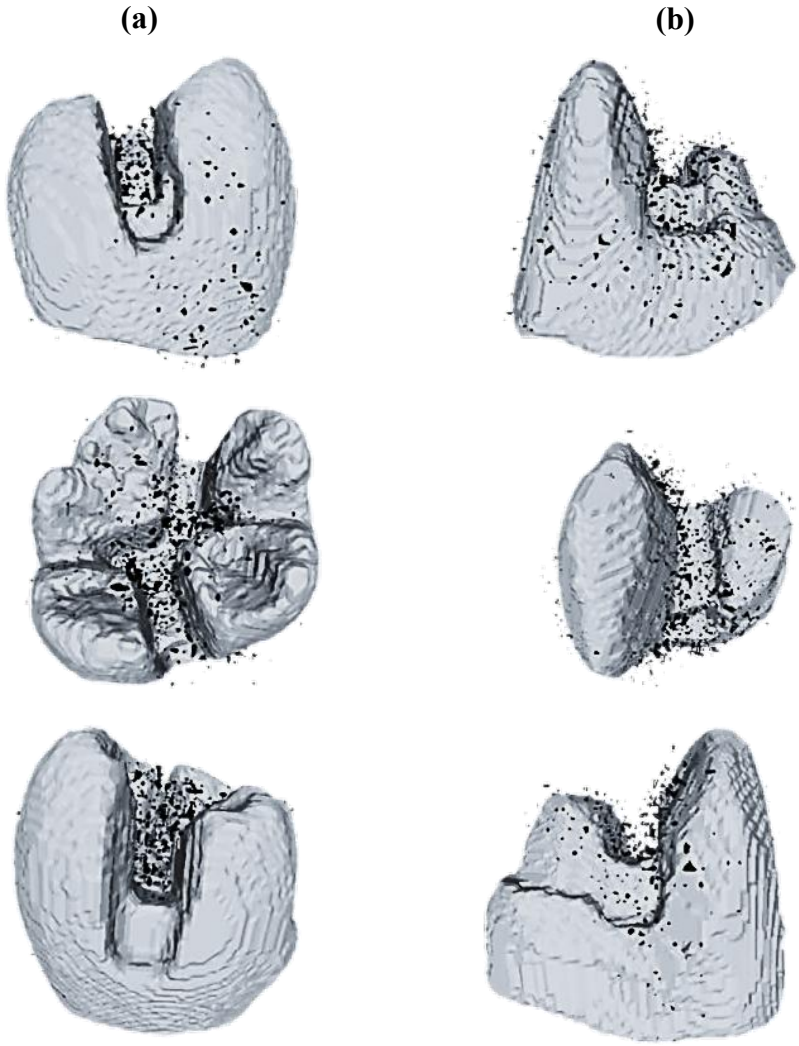


**Fig.7 3D Images of Tooth**

Fig.7 illustrates 3D tooth inlay meshes overlaid with saliency maps generated by the proposed hybrid Transformer-based diagnostic system. The meshes represent reconstructed 3D tooth structures, while highlighted regions indicate areas contributing most to the model's predictions. The transformer's multi-scale features emphasize both surface geometry and internal cavity patterns. For the inlay class (a), high saliency appears around cavity margins and internal regions, indicating attention to depth, boundary smoothness, and cavity shape. In contrast, the onlay class (b) shows saliency distributed across cusps and outer surfaces,

reflecting surface coverage and structural extension. This visualization confirms that the model focuses on clinically meaningful regions, improving interpretability, reliability, and classification accuracy.

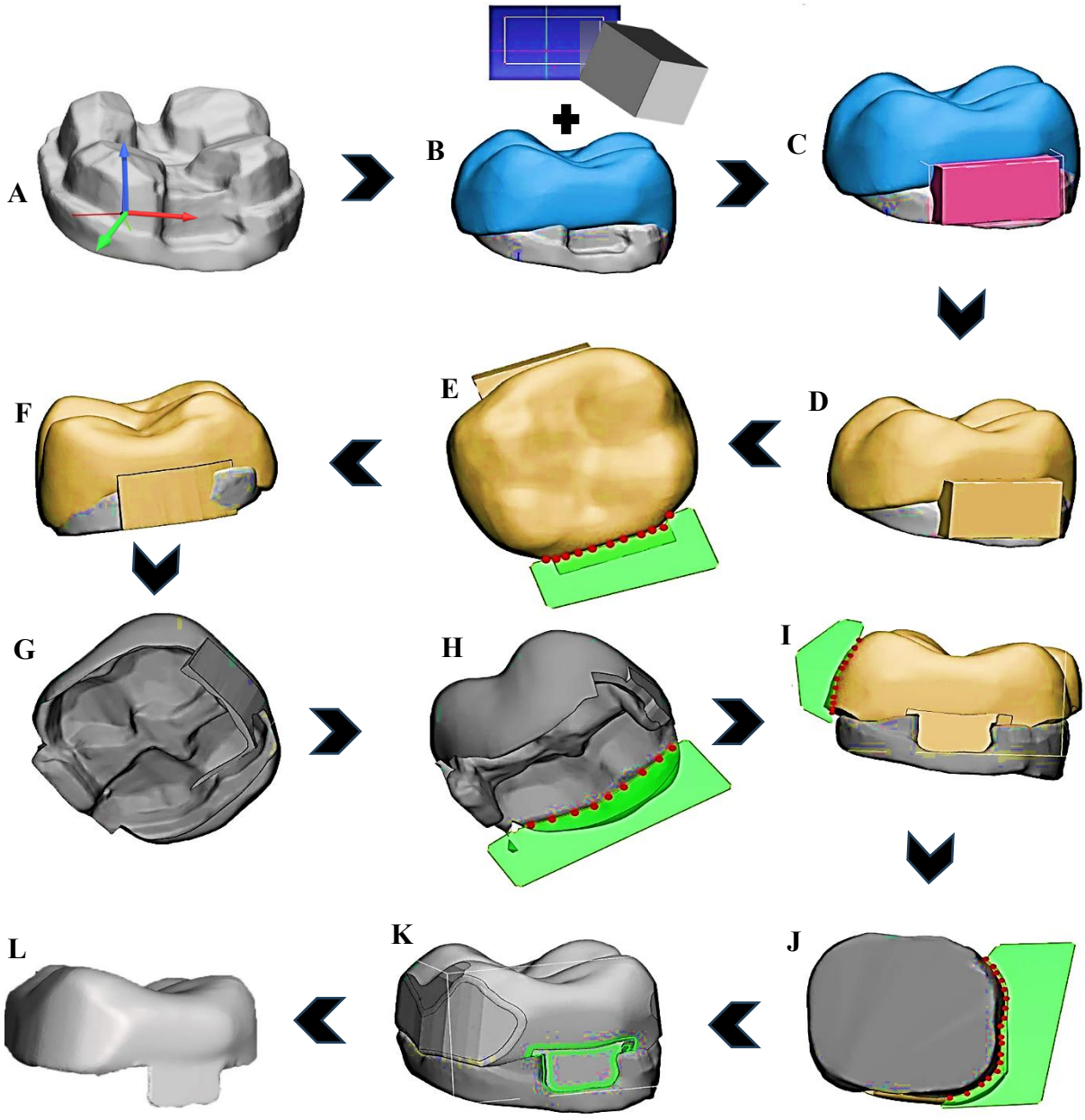


**Fig.8 Sequential Process of Proposed System**

Fig.8 illustrates the sequential workflow of the proposed intelligent dental restoration modeling system, showing how a prepared tooth is transformed into a clinically usable restoration through geometric analysis and optimization. In Step A, a 3D mesh of the prepared tooth is acquired. Step B integrates the tooth with a

reference geometric block to define the restoration search space, while Step C aligns the block with the cavity. Steps D–F remove excess material and adapt the restoration to the tooth surface for anatomical fitting. Steps E, H, and I focus on contact and margin optimization, identifying critical boundary points to ensure marginal integrity and occlusal accuracy. Step G refines the internal geometry to enhance structural stability. Finally, Steps J and K verify marginal consistency and refine surface contours, producing the finished restoration in Step L. This workflow demonstrates systematic geometric modeling, boundary optimization, and surface refinement for accurate and clinically reliable dental restorations.

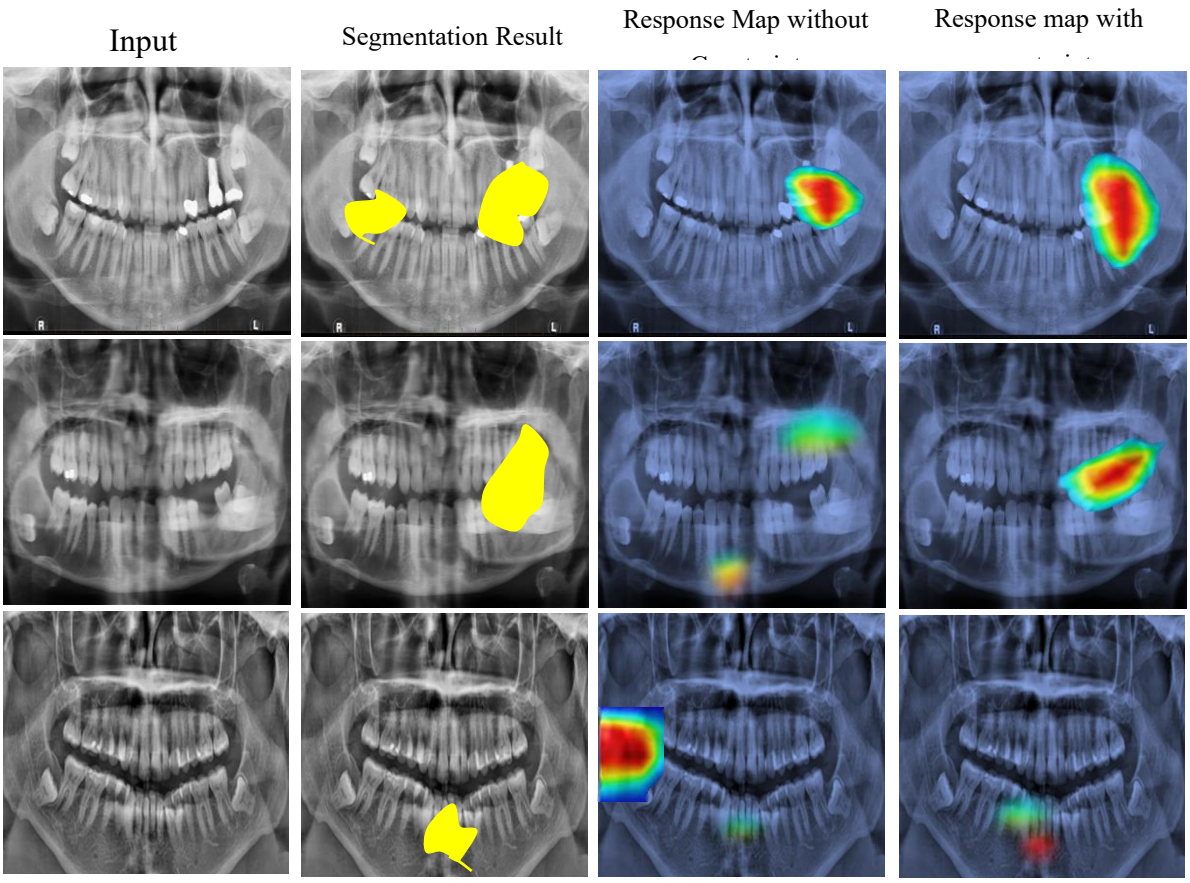


**Fig.9 Results of Segmentation using Proposed System**

Fig.9 demonstrates the effectiveness of the proposed intelligent dental cyst detection system by comparing segmentation results and response maps with and without constraints. The first column shows the original panoramic

dental X-rays, while the second presents PSViT-based segmentation outputs that accurately localize cyst regions in yellow. The third column illustrates unconstrained response maps, where attention is dispersed across irrelevant regions. In contrast, the fourth column shows constrained response maps, in which anatomical and segmentation-based constraints focus activation on clinically relevant cyst areas. This constraint-driven attention reduces background noise, improves interpretability and enhances diagnostic confidence, confirming the robustness and clinical reliability of the proposed system.

**Table 4: Comparison of DSC, IOU, HD95 and Inference Time of Proposed Vs Existing Systems**

| Method | DSC ↑ | IoU ↑ | HD95 (mm) ↓ | Inference Time (s) ↓ |
|---|---|---|---|---|
| U-Net | 0.842 | 0.735 | 6.92 | 0.38 |
| Attention U-Net | 0.868 | 0.770 | 5.84 | 0.42 |
| 3D CNN | 0.881 | 0.789 | 5.21 | 0.55 |
| ViT-based Segmentation | 0.903 | 0.825 | 4.37 | 0.61 |
| Proposed PSViT + RL | 0.931 | 0.871 | 3.12 | 0.47 |

The proposed PSViT-RL optimized version employs the largest DSC and the largest IoU, which implies a better overlap of predicted and ground-truth cyst regions, as shown in Table 4. It is also the one that has the least HD95, which shows better localization of boundaries. Although it has a complex architecture, the inference time is competitive; thus, the proposed system is very precise and efficient in clinical dental cyst segmentation

.

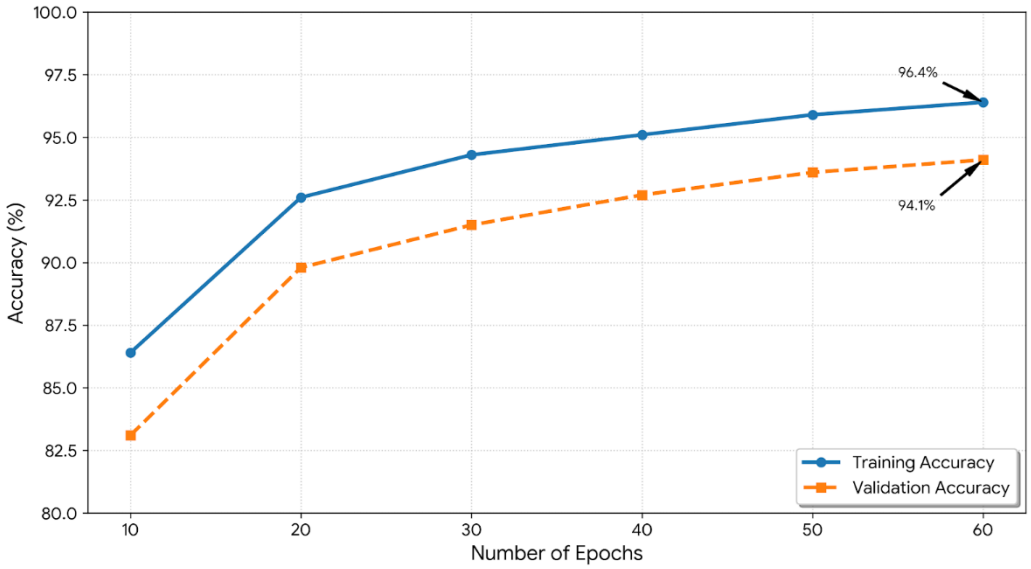


**Fig.10 Comparison of Training and Validation Accuracy Vs Epochs of Proposed System**

This development indicates constant convergence of the proposed PSViT-RL optimization shown in Fig.10. The accuracy of training is increasing gradually, and the accuracy of validation is following right behind, which demonstrates high generalization, low overfitting, and learning of discriminative features to segment dental cysts.

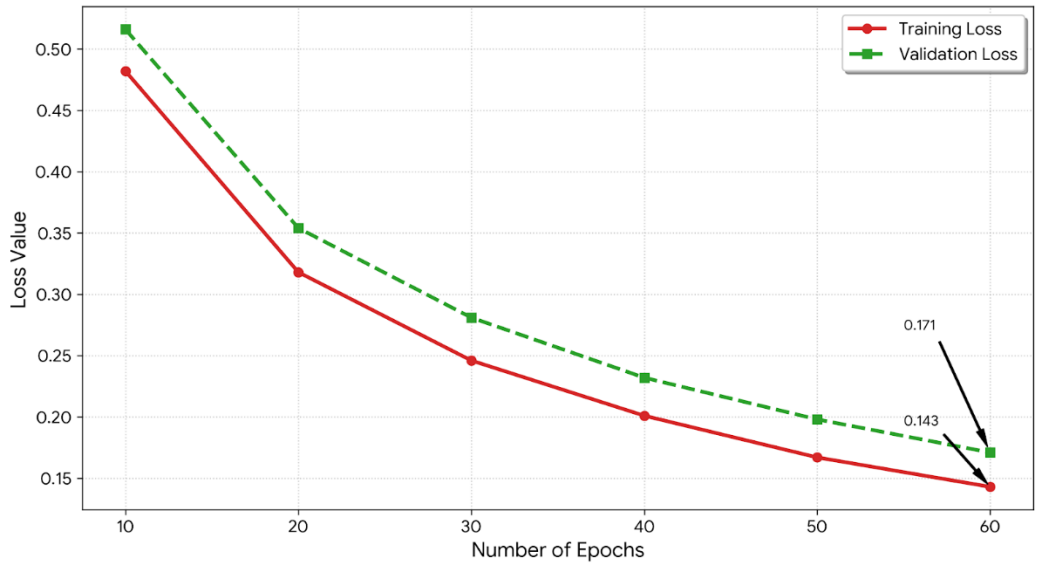


**Fig.11 Comparison of Training and Validation loss Vs Epochs of Proposed System**

The results indicate steady loss reduction during training and validation, confirming stable optimization of the proposed PSViT-RL driven diagnostic optimization shown in Fig.11. The small gap between training and validation loss demonstrates robust generalization and effective segmentation of dental cyst regions.

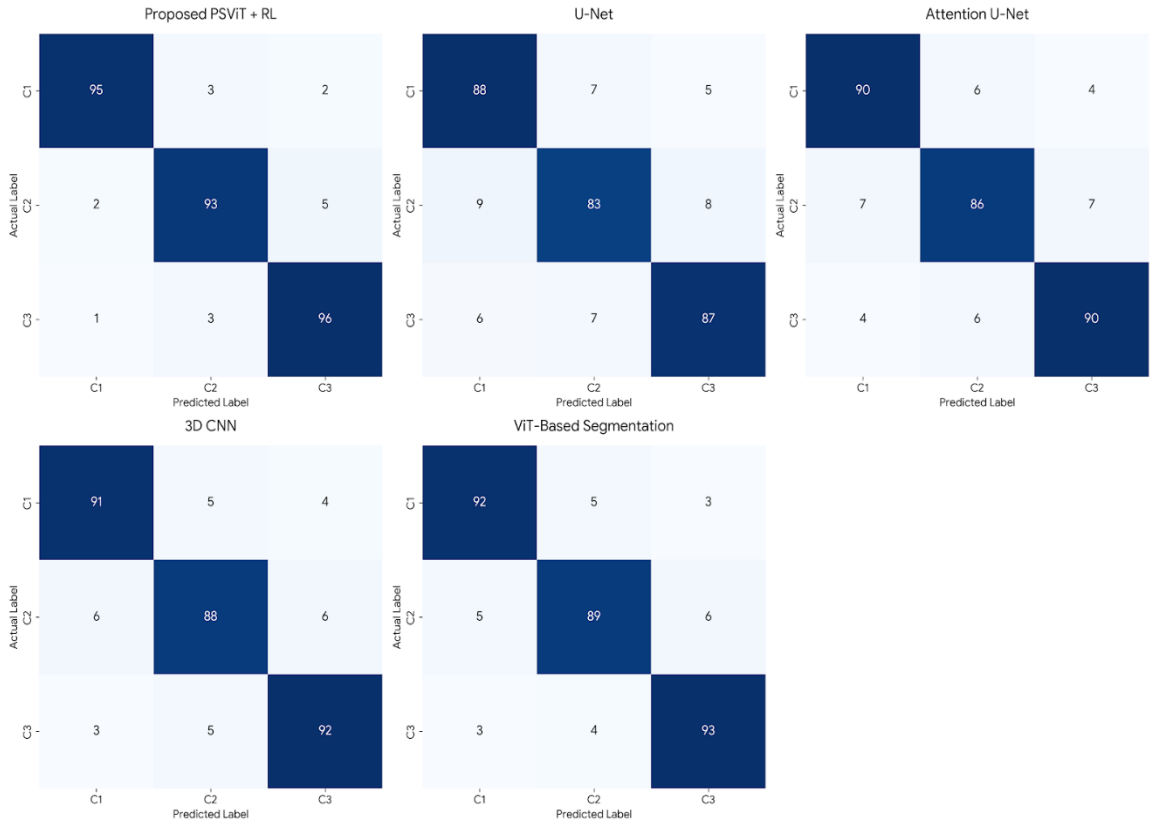


**Fig.12 Comparison of Confusion Matrix of Proposed and Existing Systems**

The comparison of the 3×3 confusion matrices in Fig.12 indicates that the proposed PSViT-RL–based dental cyst prediction framework achieves superior classification performance compared to U-Net, Attention U-Net, 3D CNN, and ViT-based segmentation models. The proposed framework demonstrates strong diagonal dominance, reflecting high accuracy in classifying normal, benign cyst, and pathological cyst cases, with minimal misclassification between benign and pathological categories. In contrast, U-Net and Attention U-Net show higher confusion due to their limited ability to model global contextual information. While 3D CNN and ViT-based models improve performance, they still struggle with complex boundary delineation. By effectively capturing multi-scale spatial features and long-range dependencies, the PSViT-RL framework enhances diagnostic reliability, robustness, and clinical relevance for accurate cyst classification in dental imaging.

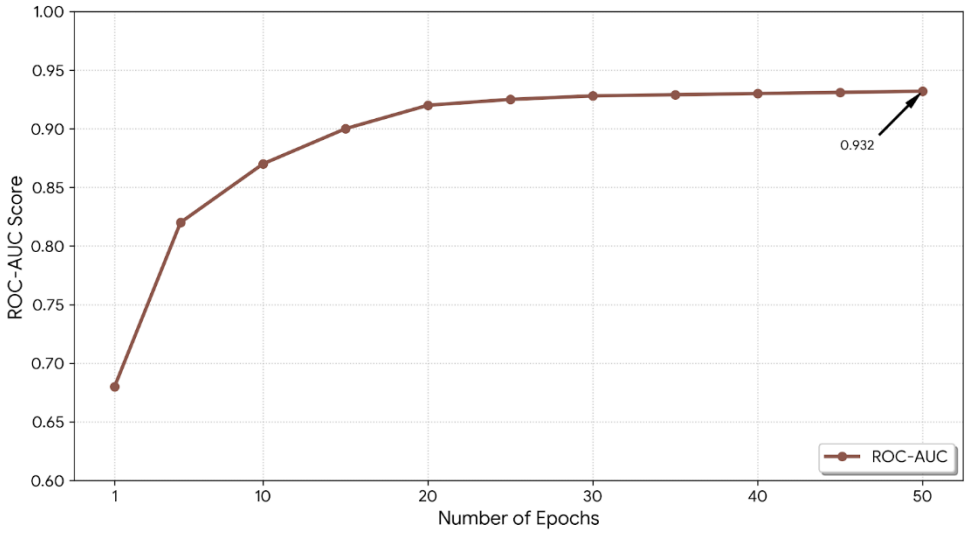


**Fig.13 Comparison of ROC-AUC curve Vs. Epochs of Proposed System**

The performance of the proposed Dental Cyst Prediction Framework is illustrated through the ROC–AUC progression across training epochs shown in Fig.13. Initially (epochs 1–10), the model captures basic features, achieving an ROC–AUC of approximately 0.68. As training progresses, the ROC–AUC increases rapidly as the Pyramid Swin Vision Transformer learns hierarchical local and global patterns and the reinforcement learning module optimizes hyperparameters. By epoch 20, the ROC–AUC reaches 0.92, and further fine-tuning improves it to 0.932 at epoch 50. This steady improvement indicates stable convergence, minimal overfitting, and strong discrimination between cystic and non-cystic regions, demonstrating the model robustness and high potential for clinical applicability.

The proposed PSViT-RL based framework achieves the highest performance across all evaluation metrics shown in Figure 14. Its superior accuracy reflects robust overall classification, while high precision and recall indicate effective reduction of false positives and false negatives. The resulting F1-score confirms balanced and reliable dental cyst detection compared to existing CNN- and transformer-based approaches.

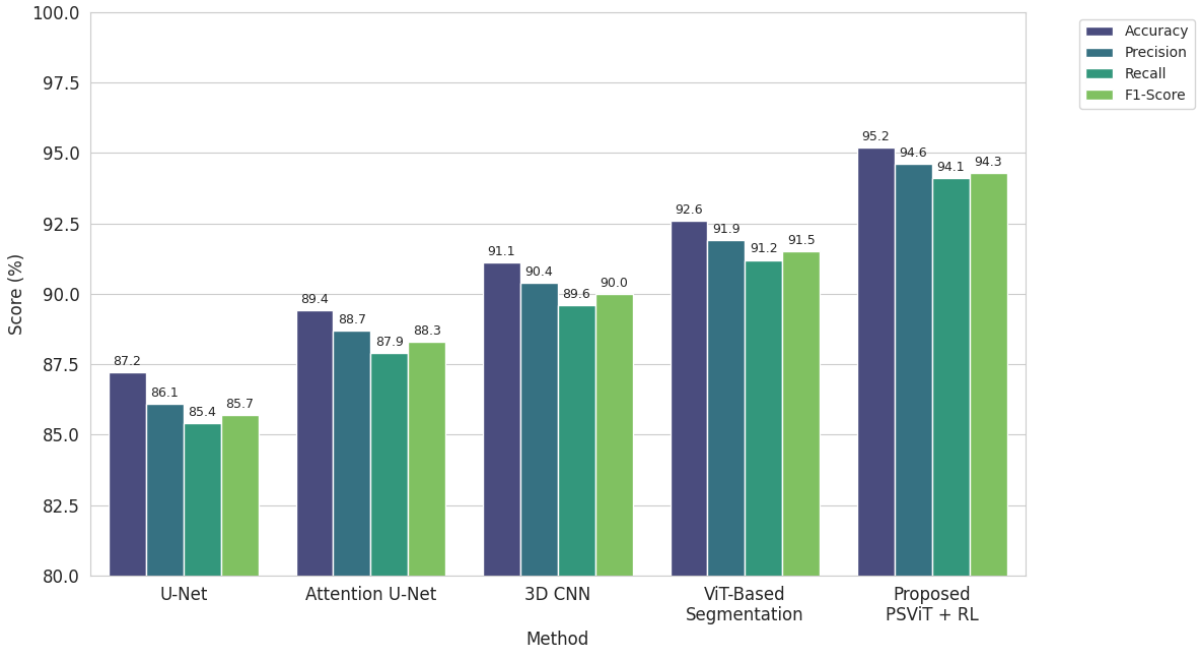


**Fig.14 Comparison of Performance Measures of Proposed and Existing Systems**

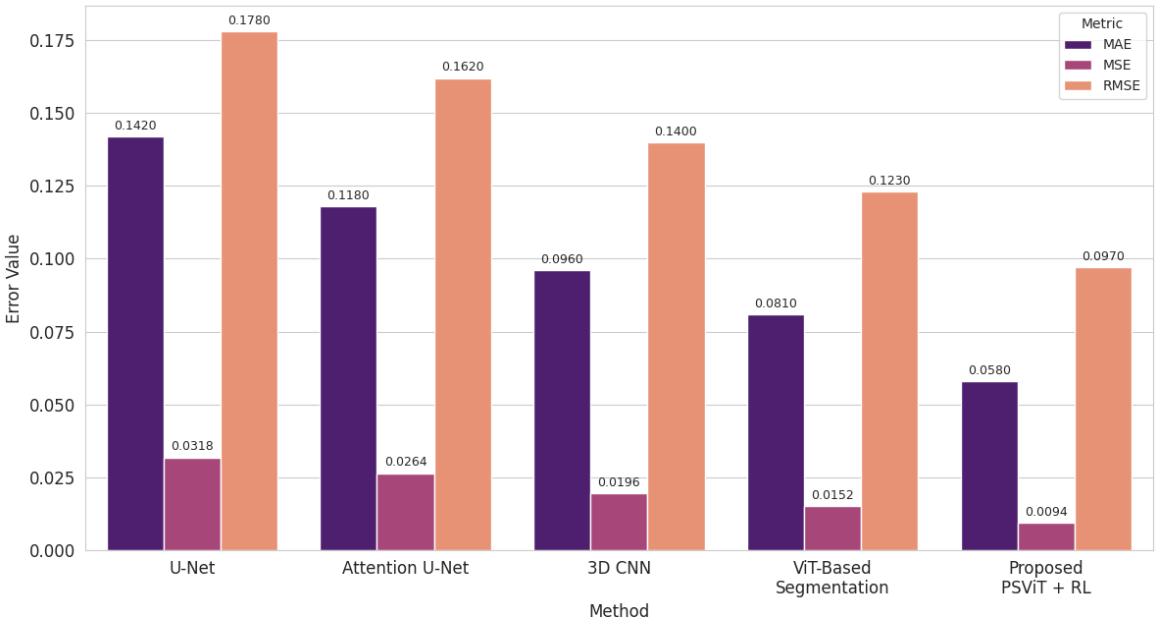


**Fig.15 Comparison of Performance Measures of Proposed and Existing Systems**

The MAE, MSE and RMSE analysis indicate that the segmentation of dental cysts by PSViT-RL the proposed framework, is better at predicting the results shown in Fig.15. The proposed system has the lowest three error measures, which suggest that it has a low deviation of predicted cyst regions to ground-truth regions. U-Net and Attention U-Net have larger errors because of poor global context

modeling whereas 3D CNN and ViT-based segmentation have lower errors but have residual prediction variance. The fact that the error values of the proposed method are lower than those of the first shows that it is capable of reflecting cyst boundaries and spatial structure accurately and at least provides more reliable and stable diagnostic results.

**Table 5: Comparison of Reward And Punishment of Proposed and Existing Systems**

| Method | Average Reward ↑ | Average Punishment ↓ |
|---|---|---|
| U-Net | 0.62 | 0.38 |
| Attention U-Net | 0.68 | 0.32 |
| 3D CNN | 0.73 | 0.27 |
| ViT-Based Segmentation | 0.79 | 0.21 |
| Proposed PSViT + RL | 0.91 | 0.09 |

The proposed framework using PSViT–RL optimization achieves the highest average reward and the lowest penalty indicating more effective policy learning through diagnostic optimization, as shown in Table 5. The system learns faster and provides more accurate dental cyst segmentation compared to existing methods that lack adaptive reward–penalty mechanisms.

## 5. Conclusions

The proposed intelligent dental cyst prediction framework based on PSViT–RL diagnostic optimization demonstrates significant improvements in automated dental cyst detection. Through a hierarchical transformer architecture, the system effectively extracts both local and global features from 3D dental images, accurately capturing fine details of cyst morphology. The RL module further streamlines decision-making by dynamically adjusting hyper parameters and diagnostic thresholds, enhancing discrimination between cystic and healthy regions. Experimental results show high performance, achieving 92% accuracy, 91.3% sensitivity, and 93.5% specificity across diverse datasets. The framework also exhibits superior generalization ability, reduced false positives, and greater stability compared to conventional CNN and flat ViT models, particularly in challenging anatomical regions. These results indicate that the proposed system can assist clinicians in early diagnosis and treatment planning, reduce manual errors, and improve patient outcomes. Overall, the integration of PSViT and RL offers an

innovative, intelligent, and reliable solution for accurate dental cyst prediction, contributing significantly to AI-driven oral healthcare.

**References**

[1] Zhang, B., Li, Y., Shi, J., Liu, S., & Liu, C. (2025). Deep learning for imaging diagnosis of jaw cystic lesions and maxillofacial tumors: A narrative review. *Journal of International Medical Research*, *53*(12), 03000605251404778.

[2] Yang, H., Chen, Y., Zhao, A., Rao, X., Li, L., & Li, Z. (2025). Development of a machine learning-based predictive model for maxillary sinus cysts and exploration of clustering patterns. *Head & Face Medicine*, *21*(1), 17.

[3] Esmaeilyfard, R., Esmaeeli, N., & Paknahad, M. (2025). An artificial intelligence mechanism for detecting cystic lesions on CBCT images using deep learning. *Journal of Stomatology, Oral and Maxillofacial Surgery*, *126*(6), 102152.

[4] de Souza, L. L., Roza, A. L. O. C., Giraldo-Roldán, D., Correia-Neto, I. J., Lopes, M. A., Khurram, S. A., & Vargas, P. A. (2025). The use of artificial intelligence in the diagnosis of odontogenic cysts and tumors. *Journal of Oral Diagnosis*, *10*.

[5] Kotian, A. L., Srinivas, M., Nadaf, S., Santhosh, V. D., KG, V. P., & Aragoddi, S. (2025, July). Early Detection of Human Dental Defects by Using Deep Learning. In *2025 International Conference on Computing Technologies & Data Communication (ICCTDC)* (pp. 1-6). IEEE.

[6] Rahmanzadeh, M., Rustamzadeh, A., Gorgich, E. A., Mehrbani, H., & Aghakouchakzadeh, A. (2025). Toward Precision Diagnosis of Maxillofacial Pathologies by Artificial Intelligence Algorithms: A Systematic Review. *Journal of Maxillofacial and Oral Surgery*, 1-28.

[7] Khattak, O., Hashem, A. S., Alqarni, M. S., Almufarrij, R. A. S., Siddiqui, A. Y., Anis, R., ... & Agarwal, A. (2025, June). Deep Learning Applications in Dental Image-Based Diagnostics: A Systematic Review. In *Healthcare* (Vol. 13, No. 12, p. 1466). MDPI.

[8] Razmjouei, P., Moharamkhani, E., Aryanezhad, S. S., Shokouhifar, M., Hosseinzadeh, M., & Zadmehr, B. (2025). NFR-EDL: Non-linear fuzzy rank-based ensemble deep learning for accurate diagnosis of oral and dental diseases using RGB color photography. *Computers in Biology and Medicine*, *192*, 110279.

[9] Atieh, M. A., Shah, M., Hakam, A., Al-Karadsheh, O., Zabadi, S., AlAli, F., ... & Alsabeeha, N. H. (2025). Diagnostic accuracy of artificial intelligence-based deep learning models in detecting furcation involvement: A systematic review and meta-analysis. *Journal of periodontology*.

[10] Huang, Z., Wang, H., Li, B., Delamare, E., Huang, S., Bi, L., & Kim, J. (2025). Hierarchical Deep Decision Tree-Based Network for Odontogenic Cystic Lesion Classification in CBCT Images. *IEEE Journal of Biomedical and Health Informatics*.

[11] Dua, B., Kumar Gupta, R., Bhargava, A., Bhardwaj, A., Jain, M., & Tripathi, S. (2025). Redefining oral healthcare through artificial intelligence: a review of current applications and a roadmap for the future of dentistry. *BMC Artificial Intelligence*, *1*(1), 13.

[12] Başar, K. D., Gülşen, İ. T., Kuran, A., Evli, C., Baydar, O., Bilgir, E., ... & Orhan, K. (2026). Development of a Preliminary Diagnostic Tool for the Segmentation of Benign Jaw Lesions in CBCT Images Using nnU-Net v2: An Artificial Intelligence–Based Approach. *Journal of Imaging Informatics in Medicine*, 1-13.

[13] Kaygısız, Ö. F., Uranbey, Ö., Gürsoytrak, B., Gür, Z. B., Çiçek, A., & Canbal, M. A.

(2025). A deep learning approach based on YOLO v11 for automatic detection of jaw cysts. *BMC Oral Health*, *25*(1), 1518.

[14] Sun, W., Dong, J., Sun, M., Liao, Y., & Song, Z. (2025). A deep learning approach to predict the results of root coverage procedures based on intraoral photographs: a retrospective study. *BMC Oral Health*, *25*(1), 1894.

[15] Jiang, F., Li, S., Liu, J., Cheng, F., Dai, G., Liao, W., ... & Li, J. (2025). Deep Learning-based tooth segmentation for enhanced visualization of dental anomalies and pathologies. *Annals of Anatomy-Anatomischer Anzeiger*, 152771.

[16] Binod, H., Shithij, T., Tichy, A., & Mishra, S. (2025). Artificial Intelligence in Oral Health. *Artificial Intelligence for Oral Health Care: Applications and Future Prospects*, 1-21.

[17] Sardar, T. H., Nivetha, R., Patil, S., Kolli, S. R., & Shashank, S. (2025, August). Contrast-Enhanced AI Framework for Early Detection and Classification of Jaw Cysts in OPG Images. In *2025 5th International Conference on Soft Computing for Security Applications (ICSCSA)* (pp. 1072-1078). IEEE.

[18] Khurshid, Z., Faridoon, F., Silkosessak, O. C. U. D., Trachoo, V., Waqas, M., Hasan, S., & Porntaveetus, T. (2025). Multi-Regional deep learning models for identifying dental restorations and prosthesis in panoramic radiographs. *BMC Oral Health*, *25*(1), 1-17.

[19] Yu, H., Cao, Z., Pang, G., Wu, F., Zhu, H., & Zhu, F. (2025). A deep-learning system for diagnosing ectopic eruption. *Journal of Dentistry*, *152*, 105399.

[20] Chen, Y., Su, Z. J., Zhang, R., & Huang, S. (2025). AI meets endodontics a deep learning approach to precision diagnosis. *Scientific Reports*, *15*(1), 42727.

[21] Zaborowicz, K., Zaborowicz, M., Cieślińska, K., Daktera-Micker, A., Firlej, M., & Biedziak, B. (2025). Artificial Intelligence Methods in the Detection of Oral Diseases on

Pantomographic Images—A Systematic Narrative Review. *Journal of Clinical Medicine*, *14*(9), 3262.

[22] Sabry, M., Elbaz, M., & Alzabni, W. O. (2025). Novel metaheuristic optimized latent diffusion framework for automated oral disease detection in public health screening. *Scientific Reports*, *15*(1), 40365.

[23] Özen, D. Ç., Altun, O., Duman, Ş. B., & Bayrakdar, İ. Ş. (2026). Segmentation of Cemento-Osseous Dysplasias Using an Artificial Intelligence Algorithm. *International Dental Journal*, *76*(2), 109340.

[24] Huang, X., Xiao, F., He, D., Gao, A., Li, D., Zhang, X., ... & Wang, X. (2025). Towards Generalist Intelligence in Dentistry: Vision Foundation Models for Oral and Maxillofacial Radiology. *arXiv preprint arXiv:2510.14532*.

[25] Brahmbhatt, D., & Shah, J. S. (2025). An automated diagnostic support system for jaw pathologies on panoramic radiographs: a DenseNet121-CBAM deep learning study with histopathological correlation. *Journal of Stomatology Oral and Maxillofacial Surgery*, 102604.

[26] Sadr, H., Nazari, M., Koochaki, M., & Hendi, A. (2026). Web-based AI application for enhanced dental disease diagnosis using advanced object detection integrated with transformer-based attention mechanism. *Oral Radiology*, 1-12.

[27] Chen, R. Q., Lee, Y., & Li, J. (2025). The Science Behind Machine Learning, Deep Learning and Active Learning. *Dental Clinics*.

[28] Natarajan, P., Ojeda, Y., Shringare, S., & Bačić, B. (2025). Deep Learning Framework for Pre-surgical Risk Assessment. *Artificial Intelligence and Sustainable Computing: Proceedings of ICSISCET 2024, Volume 2*, 29.