Ibn Al-Haitham Journal for Pure and Applied Sciences
Journal homepage: jih.uobaghdad.edu.iq
PISSN: 1609-4042, EISSN: 2521-3407
IHJPAS. 2025, 38(3)

Detection and Identification of Dental Caries Using Segmentation Techniques

Noor A. Ibraheem[1]*

[1] Department of Computer Science, College of Science for Women, University of Baghdad, Baghdad, Iraq.

*Corresponding Author.

Received: 7 January 2025 Accepted:16 April 2025 Published: 20 July 2025

doi.org/10.30526/38.3.4099

Abstract

Dental caries, also named tooth decay, is a major issue for oral health and is caused by bacteria in dental plaque. Detecting caries early on is essential for preventing further damage. Because caries are often small, they can lead to unnecessary treatments or missed diagnoses. This study tackles the challenge of spotting dental caries using color image analysis. We tested both traditional methods—like Quickshift, Simple Linear Iterative Clustering (SLIC) superpixels, and k-means clustering—combined in the Multi-Step Segmentation with K-Means (MSS-KM) approach, as well as a more advanced deep learning method using YOLOv12 for segmentation. The evaluation of the performance of these methods based on accuracy, precision, recall, mean average precision (mAP), and F1-score. The results were impressive, showing that YOLOv12 clearly outperforms MSS-KM in terms of accuracy. YOLOv12 achieved an accuracy of 98.12%, while MSS-KM was at 96.79%. In addition to accuracy, YOLOv12 had excellent precision (99.6%), recall (98.1%), and an F1-score of 0.99, while MSS-KM came in at 88.5% for precision, 91% for recall, and an F1-score of 89.1%. YOLOv12 also had a mAP of 99.5%, compared to MSS-KM's 99.3%. These results clearly show that YOLOv12 is more accurate and reliable for detecting dental caries than MSS-KM. While the MSS-KM method still has value, particularly for traditional segmentation techniques, the model showed strong potential for practical use in clinical settings. The consistent training setup contributed to stable performance, while the comparison with traditional methods highlighted how modern deep learning approaches can significantly enhance diagnostic accuracy. These results not only support the use of YOLOv12 for early caries detection but also suggest that such AI models could become valuable tools in improving patient outcomes and reducing unnecessary treatments.

Keywords: Deep learning, YOLO v12, teeth detection, Dental caries, Segmentation, K-Means clustering.

1. Introduction

The World Health Organization (WHO) describes tooth decay as the breakdown of a tooth's outer surface, caused by acids produced when bacteria feed on sugars in the mouth (1, 2).

Dental caries compose a condition that significantly impacts oral health, which can cause degradation of enamel and dentin due to the action of bacteria present in dental plaque. If left untreated, the condition may spread into the dental pulp, which carries nerves and blood vessels inside the tooth, causing inflammation and possibly even tooth loss (3-5).

The most prevalent oral health issue brought through bacteria interacting with the mineral enamel is dental caries. Insufficient fluoride and oral component are associated risks for tooth decay. Nevertheless, the advancement of cavities can be affected by additional variables, including attitudes, food choices, sanitation, societal position, quality of lifestyle, and socio-demographic characteristics (6, 7).

In order to decide how the treatment will be carried out, it is crucial to obtain an accurate clinical diagnosis in dentistry (8). Image segmentation considered as a significant part in different image processing fields. Dental images fall into one of two categories: extraoral or intraoral. Extraoral radiographs called panoramic X-ray images are utilized to identify dental issues in the maxilla and jaw (8), in this work intraoral images are used.

Two different approaches to dental image segmentation are explored. The first approach utilizes traditional segmentation methods, including Quickshift, Simple Linear Iterative Clustering (SLIC) superpixels, and k-means clustering, which named the MSS-KM (Multi-Step Segmentation with K-Means) method. These techniques are widely used in image segmentation to divide an image into more usable, smaller parts that make it simpler to identify important details such as dental caries.

The second approach employs the advanced YOLO (You Only Look Once) algorithm, which has gained recognition for its accuracy and efficiency in the segmentation process, making it effective and accurate for detecting dental issues. Both methods aim to enhance the accuracy of dental caries detection, with providing clear advantages.

The development of deep learning technologies has enhanced medical image analysis by providing strong tools for disease detection and diagnosis from medical images (9). Deep learning algorithms demonstrate significant effectiveness in automating dental caries detection through the analysis of radiographs and dental images. The U-Net architecture is widely adopted for medical imaging segmentation tasks (10), at the pixel level but does not perform well for object detection applications like dental caries detection (11). YOLO algorithm stands out as a state-of-the-art deep learning framework for real-time object detection which excels at identifying dental caries locations within dental imaging (12). This research uses YOLOv12 to improve the precision of caries detection while providing a dependable substitute for conventional segmentation techniques.

The main contribution of this work is the development of a system for detecting dental caries by applying two different segmentation techniques: a traditional Multi-Step Segmentation (MSS) method and the deep learning-based YOLOv12 model.

The rest of the paper is organized as follows: section 2 introduces the related works. Section 3 explains system material and methods. Results and discussion are presented in section 4. Section 5 concludes the work.

1.1. Related Works

Numerous studies have explored the classification and segmentation of dental caries using various imaging modalities, with the aim of improving diagnostic precision. This section provides an overview of relevant literature, highlighting key methods and findings; and presents a comparative summary of these studies in tabular form to facilitate direct comparison of their approaches, datasets, and performance metrics.

Studies using several imaging modalities have been conducted with the goal of classifying and segmenting caries. The technique of separating each object in an images from the remainder of it such that the objects do not conflict with the surroundings is known as image segmentation. The purpose of segmenting caries images is to distinguish tooth decay in the images. In this section, a brief survey on some conventional and deep learning techniques. Because dental segmentation makes it more convenient for dental professionals to conduct assessments and determine the ideal method of action for their patients, it has become increasingly important in the field of dentistry.

Researchers in (6) compare four segmentation algorithms (U-Net, DCU-Net, DoubleUNet and Nano-Net) that are well-known in the medical literature on segmentation and assess their performance against the most advanced in dental segmentation as it stands in panoramic radiography. The algorithms examined in 1500 images. They discovered that DoubleU-Net model is outstanding among other models (6). Researchers in (13) suggested multiple levels structured segmentation system using simple liner iterative clustering (LI-SLIC) method, and probability distribution similarity method to limit the amount of superpixels and get rid of over-segmentation, neighboring superpixels that belong to identical object (13). While researchers in (14) contrasted segmentation based on regions Effectiveness uses four algorithms which are; Felzenszwalb and Huttenlocher (FH), Compact Watershed (CW), Quick Shift (QS), and SLIC, with two different datasets: variation of information set (VOI) and adapted rand error set (ARE) to study real-world RS images with region of interest (ROIs). The experiments demonstrated that for both images with varying ROI difficulties, the SLIC outperformed the other algorithms in terms of outcomes (14). A comprehensive study for around 150 CNN-based techniques have been created in the past decade for semantic segmentation (15-16), two datasets are used and each includes scenarios in both two-dimensional and three-dimensional image and video frames, including scenarios in both two-dimensional and three-dimensional image and video frames, including in general, the inside, outside, and street-based circumstances. Furthermore in reference (17), three primary reviews of image segmentation phases deep learning-based segmentation using semantics, interactive segmentation, and traditional segmentation. Statistical methods are used as well, as in references (18-19).

Multilayer neural network models are used in deep learning, a type of machine learning, for a range of applications, such as image, video, and audio processing. By mechanically separating characteristics from plain data symbols rather than learning using rules, deep learning can gain these characteristics concurrently, in contrast to typical machine learning techniques (20)(45). For deep learning methods, researchers in (21), developed a framework named DENTECT that create an enamel number depending on the FDI assessment on panoramic X-ray images while simultaneously recognizing five different dental therapies procedures (21). Convolution Neural Network CNN is applied in (8) to determine whether teeth are present or absent on the radiographs. Additionally in reference (22), a CNN s applied for segmentation and classification process on panoramic X-Ray images, every tooth is split and assigned a specific value (23). A Pilot study is presented in (24), in which MobileNet V2 is used. In (17), researchers used Multiple Label U-Net for image segmentation by utilizing augmented data. While in reference (25), the researchers proposed a Multiscale Residual Dilated U-Net (MSRD-UNet) for Medical Image Segmentation, three composite datasets were used: nuclei, lesions of the skin, and polyps, the suggested system applied CNNs and U-Net network. While in (26) authors proposed deep neural networks, and the pseudo edge-region was obtained to acquire preliminary contour for every tooth area.

Bayati et al., in (12) proposed an advanced AI-driven method for detecting interproximal caries in bitewing radiographs using the YOLOv8 model. The study demonstrates the power of YOLOv8 in identifying caries in dental radiographs, achieving high precision and recall rates for interproximal caries detection, which is a challenging task in dental imaging. The results shows that deep learning with advanced YOLO models, is highly effective for automated dental diagnostics, in detecting subtle carious lesions in bitewing radiographs (12).
Tareq et al. (27) developed a hybrid YOLO ensemble model combined with transfer learning for visual diagnostics of dental caries. The method uses non-standardized photographs of dental conditions, highlighting the priority of deep learning to automate the detection of dental caries. The hybrid approach leverages both YOLO's object detection capabilities and the generalization power of transfer learning, introducing promising results for caries detection in non-X-ray images (27).
Beser et al., in (28), explored the use of YOLOv5 in pediatric panoramic radiographs for tooth detection and segmentation in mixed dentition. The study shows the effectiveness of YOLOv5 in detecting and segmenting teeth, which is essential for various dental procedures, including caries detection. The results demonstrate that YOLOv5 provides a reliable and efficient solution for tooth detection, further more proving YOLO's versatility and precision in dental image analysis (28). A survey on various deep learning methodologies for objects recognition, segmentation, and identification are introduced in (29). An extensive analysis of deep learning methods for determining dental conditions, such as abnormalities are presented in (20).

1.2. Comparative Discussion with Previous Work

To better understand the positioning and impact of this study, a comparison with previous research efforts on dental caries detection is provided. Various studies have explored different machine learning (44) and image processing techniques, each using distinct datasets and evaluation strategies. **Tables (1)** and **(2)** below summarizes key aspects of these works, including the type of image data used, variables detected, applied methodologies, dataset sources, training and evaluation strategies, and the resulting accuracy. This comparative overview highlights methodological trends and variations across studies, offering context for the results achieved in the present work.

Table 1. Comparative Analysis of Related Work on Dental Caries Detection

Reference, Author, Year	Type of image data used	Variable Detected	Applied techniques
(5) Luiz Guilherme Kasputis Zanini *et al.*, 2024.	Cone Beam Computed Tomography, (3D) images	teeth caries lesions	Six methods: Naïve Bayes (NB), RF, SVM, K-Nearest Neighbors (KNN), logistic regression (LR), and XGBoost (XG).
(8) María Prados-Privado *et al*., 2021.	panoramic images	Teeth caries	ResNet Atrous Convolution
(12) Bayati, M. et al, 2025	Interproximal caries detection in bitewing radiographs.	caries detection	YOLOv8
(17) Rini Widyaningrum *et al.*, 2022.	panoramic radiographs	Periodontitis	Multi-Label U-Net, and Mask R-CNN models
(21) Atıf Emre Yüksel *et al*., 2021.	Panoramic X‑rays	Root canal therapy (RCT), lesion treatment, dental	deep learning method

Reference, Author, Year	Type of image data used	Variable Detected	Applied techniques
		fillings surgical the extraction process, and ancient extraction.	
(22) Mircea Paul Muresan *et al*., 2020.	panoramic X-Ray images	Teeth caries	CNN
(24) Shankeeth Vinayahalingam *et al*., 2021.	Panoramic radiographs	Teeth, carious lesions in the third molars of the maxilla and mandible.	CNN MobileNet V2.
(26) Seongeun Kim *et al*., 2024.	general optical image	Teeth caries	deep neural network
(30) Dr.Riddhi Chawla et al., 2022.	radiographic images	dental caries	CNN
(31) Shuaa S. Alharbi *et al*., 2023.	X-ray radiography	Dental caries lesions.	U-Net, U-Net++ and U-Net3+
(32) Paras Tripathi *et al.*, 2019.	Radiography, X-ray images	Dental caries detection	Genetic algorithm bases BPNN
(33) Y. Jusman *et al*., 2021.	radiographic images of 4 dental caries clusters	Dental caries detection	SVM, and K-Nearest Neighbors (KNN).
(34) Aayush J. *et al*., 2023	intraoral images	dental cavities deteciton	YOLOv5, and Faster R-CNN
Suggested systems	intraoral images	Teeth caries	YOLO12, MSS-KM approach

Table 2. Benchmarking Previous Approaches for Caries Detection

Reference, Author, Year	Data set	Training/Evaluating	Accuracy Results
(5) Luiz Guilherme Kasputis Zanini *et al.*, 2024.	four databases: Origin, PCA, Chi, RF.	66 images as ICDAS 0, 32 images as ICDAS 1, 50 images as ICDAS 2, 151 images as ICDAS 3, 194 images as ICDAS 4.	86.20%
(8) María Prados-Privado *et al*., 2021.	8000 panoramic radiographies.	According to experts, after refining, 236 of 8-bits, 68 of 12-bits.	99.24%
(12) Bayati *et al*, 2025	1506 images	80% training 10%, testing 10%, validation	N/A
(17) Rini Widyaningrum,*et al.*, 2022.	1100 original and augmented images.	75%Training 25% testing.	95%
(21) Atıf Emre Yüksel *et al*., 2021.	1005 X-ray images.	85% training 15% testing	96.4%
(22) Mircea Paul Muresan *et al*., 2020.	1000 images.	70% training 10% cross-validation 20% testing.	89%
(24) Shankeeth Vinayahalingam *et al.*, 2021.	500 images.	320 for training, 80 for validation 100 for testing	87%
(26) Seongeun Kim *et al*., 2024.	Set consists of 46 images with different references.	All 46 images for training	N/A
(30) Dr.Riddhi Chawla *et al.*, 2022.	10,000 Datasets including radiographic dental cavities images.	5000 training 5000 testing	99.12%
(31) Shuaa S. Alharbi *et al*., 2023.	1500 panoramic X-ray images.	60% training 40% validation	95%

Reference, Author, Year	Data set	Training/Evaluating	Accuracy Results
(32) Paras Tripathi *et al.*, 2019.	800 images	N/A	95.42%
(33) Y. Jusman *et al.*, 2021.	264 images	198 training 66 testing.	100%
(34) Aayush *et al.*, 2023	300 images	80% training 20%, testing	YOLOv5 = 75%, and Faster R-CNN= 80%
Suggested systems	5725 images	74% training 21% testing 5% validation	98.12%

As shown in the comparison, many existing approaches rely on smaller or limited datasets, and the methods applied vary significantly, ranging from classical machine learning algorithms to early deep learning models. While some systems achieved reasonable accuracy, the majority lack consistency in evaluation protocols and use of standardized datasets. In contrast, the proposed approach in this study utilizes a more structured training setup and leverages a modern object detection model, YOLOv12, which demonstrates superior accuracy and robustness in caries detection. This reflects an advancement in both technique and reliability, addressing key limitations identified in earlier research.

2. System Material and Methods

This paper compares two different approaches for detecting dental caries. The first approach, MSS-KM (Multi-Step Segmentation with K-Means), using traditional image segmentation methods like Quickshift (35), SLIC (36-38), and K-Means clutering (39-41). Since these methods are well-established and effective, yet still considered more traditional as advanced as newer techniques. The second approach, on the other hand, takes advantage of the powerful YOLOv12 algorithm (42), which is a cutting-edge deep learning model known for its speed and accuracy in detecting objects. Figure 1 illustrates the core components of the proposed dental detection system, shows how these two methods offer unique insights into the comparison of traditional versus modern segmentation techniques.

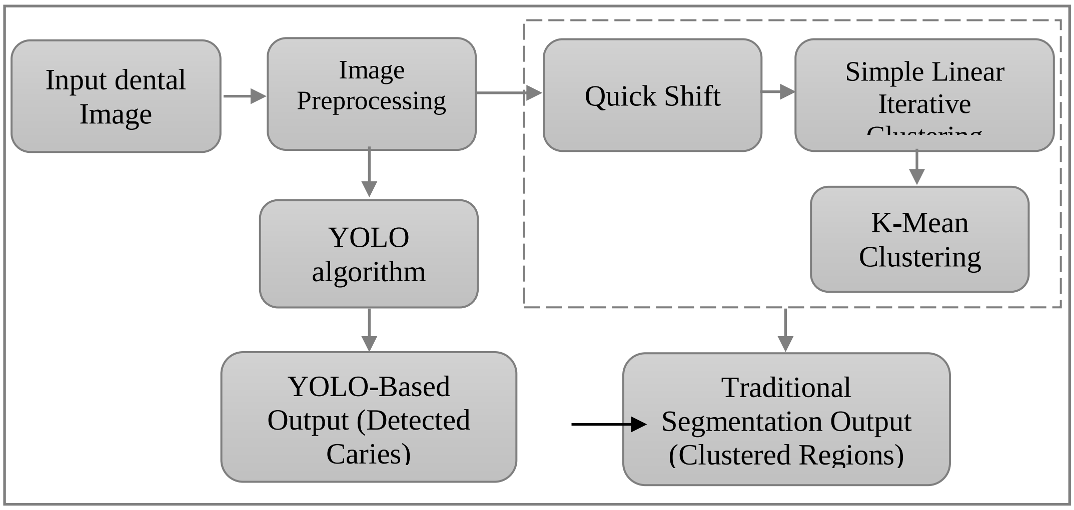

Figure 1. Flowchart summary of the proposed dental method.

2.1. Dental Images Segmentation

Grouping small, meaningful regions in an image known as superpixels has become increasingly popular in computer vision. This approach helps speed up the final stages of image processing and is widely used in various applications.

2.2. Dataset Input

This study utilized publicly available datasets from platforms like Kaggle and Roboflow, which contain annotated dental images for segmentation and classification purposes. The datasets cover a broad spectrum of cases, including different imaging conditions, tooth types, and stages of dental caries, enabling robust evaluation of model performance under clinically diverse scenarios. While not custom-collected for this research, these datasets are highly suitable for training and testing segmentation models, thanks to their pixel-level annotations and balanced representation of both healthy and carious teeth samples.

2.2.1 Dataset Overview

The dataset consists of 5725dental images, each of which is annotated with ground truth segmentation masks. The images are in JPEG format and vary in resolution, with the majority being 640x640 pixels. The dataset contains both healthy teeth and images of teeth with varying stages of dental caries.

2.2.2. Dataset Characteristics

The Kaggle dataset utilized in this study adequately meets the specific requirements of the research, and Roboflow platform which offering image quality and content appropriate for accurate detection of tooth caries. The data link is available at the Kaggle webpage (46).

The dental condition dataset is a diverse collection of oral images designed to support research and model development in the field of dental diagnostics. It includes a variety of conditions such as calculus, gingivitis, tooth discoloration, ulcers, hypodontia, and more. For the purpose of this study, only the images related to dental caries were selected. This subset allowed the focus to remain specifically on training and evaluating models for the detection and classification of carious lesions.

The dataset comprises a diverse and well-annotated collection of dental images, categorized based on specific oral conditions. It includes examples of caries, calculus, and hypodontia cases.

The images in the dataset were collected from multiple hospitals and authoritative dental websites, ensuring both clinical relevance and diversity. To enhance the dataset's robustness and improve the suggested model, various data augmentation techniques were applied during preprocessing such as rotation, flipping, scaling, and noise addition.

By supporting various annotation formats, it reduces the time spent on manual data conversion, and collecting which focuses on model training and experimentation instead of repetitive tasks.

The employed experiments are conducted on a dataset, containing images of dental decay, the dataset used in this study comprises 5725 images of tooth caries, which were partitioned into training, testing, and validation sets.

The images obtained were stored digitally in JPEG image format, The database only includes permanent teeth caries and has no patient information, including gender and age, The images excluded are those that have poor quality, high-quality distortion, and overlapping proximal surfaces due to anatomic arrangements that maintain diagnostic accuracy.

2.3. Image Pre-processing

To distinguish dental caries, pre-processing methods is the underlying step for preparing the dental images like resizing, noise removal, normalization, histogram equalization, which can be prompted by both interior and outside sources, have a critical negative influence on the input image (7). The quality of segmentation results improved as preprocessing steps reduced noise and enhanced contrast while standardizing image size before segmentation.

These methods are listed below:

1. Resizing the Image: The uploaded image undergoes resizing to fit the dimensions. The resizing step guarantees that the image will remain the same size for evaluation purposes.
2. Noise Removal: This step includes noise removal through Gaussian blur to eliminate unwanted noise from the image.
3. Normalizing the Image: The pixel values of the image are divided by 255 to normalize them to a range between 0 and 1. Normalization of image data results in standardized data which becomes appropriate for segmentation and clustering tasks.
4. Histogram Equalization: The application of histogram equalization enhances image contrast which subsequently improves image detail visibility for segmentation. In medical image processing this technique proves valuable because contrast enhancement helps to identify subtle variations.

2.4. Quick Shift Algorithm

Quick Shift is a kernel-based mode-seeking algorithm related to Mean Shift. It estimates the density of a dataset using the Parzen window technique. Given a set of N data points x_1, x_2, ..., xn, the algorithm computes a density estimate for each point without requiring gradient calculations or quadratic lower bounds (43). Instead, it shifts each pixel x_i toward its nearest neighbor with a higher density, effectively grouping similar regions together. This makes Quick Shift a powerful tool for image segmentation by identifying clusters based on density variations (35).

2.5. Simple Linear Iterative Clustering (SLIC) Superpixels

A superpixel is a group of adjacent pixels that share similar properties (e.g., color, texture), making them useful for image segmentation (37). Superpixel methods provide an effetive way to represent images by reducing complexity while preserving essential local features. Compared to pixel-by-pixel color quantization techniques, superpixel approaches offer a more structured and meaningful representation of image regions (37). The survey by Sasmal and Dhal explores the integration of superpixel techniques with clustering algorithms for image segmentation. It provides an overview of superpixel generation methods and delves into partitional clustering techniques, highlighting their challenges and solutions. The study reviews existing literature on combining superpixels with clustering for various segmentation tasks. Additionally, it presents a comparative analysis using oral pathology and leaf images to demonstrate the effectiveness of these combined approaches (38).

Due to the convenience of utilize, conformance of boundaries, fast analysis, and efficiency; in data storage; SLIC has been shown to surpass another the most recent superpixel approaches. These findings were supported by the present investigation. The required number of identically sized superpixels to be produced is the only parameter (k) in SLIC. It is possible to set up a second parameter m that regulates how compact the superpixels are. Using a regular grid with S pixels between each group center, the method first creates k initial cluster centers. The cluster focuses are adjusted in a 3×3 neighborhood to the least contrast point in order to prevent the seed from being placed on a border or a noisy pixel. The initial stage in an iterative process is allocating each image pixel to the closest cluster center (30)(38).

SLIC Superpixel Segmentation Process (41):

1. Initializing clustering center: The SLIC algorithm starts by dividing the image into a set number of segments based on user input. If an image contains N pixels and is split into K segments, each superpixel will roughly contain N/K pixels. The algorithm places initial cluster centers at the middle of each segment and assigns a unique label to each one.

2. Calculating Distance between pixels and cluster centers: Each pixel is evaluated based on how similar its color is to the cluster center and how close it is in spatial distance**.** The goal is to ensure accurate process while preventing superpixels from forming along sharp edges. To achieve this, the algorithm adjusts the clustering process based on the smallest measured distance to the cluster center.

The following equations define how the algorithm measures the relationship between each pixel and its cluster center, using distance metric (D) for ensuring a well-structured segmentation.

$$d_{xy} = \sqrt{(x_k - x_t)^2 + (y_k - y_t)^2} \tag{1}$$

$$d_{lab} = \sqrt{(l_k - l_t)^2 + (a_k - a_t)^2 + (b_k - b_t)^2} \tag{2}$$

$$D = d_{lab} + \frac{m}{H}\, d_{xy} \tag{3}$$

d_{lab} for the color variance among pixel points, d_{xy} for the distance in pixels among pixel points, and H for the percentage of color value and spatial data used in the similarity assessment. The more similar two pixels are, greater the D value.

2.6. Model selection and utilization of YOLOv12

Choosing the right model architecture plays a key role in achieving accurate image segmentation. A well-designed model helps clearly define object boundaries while minimizing false detections, ensuring efficiency within the system's constraints. These factors are important for producing reliable segmentation results in real-world applications. In this study, YOLOv12, the latest iteration of the YOLO series, was released in February 2025 by researchers from the University at Buffalo, SUNY, and the University of Chinese Academy of Sciences (42). For real-time object recognition tasks, this approach presents an attention-based architectural framework that improves detection accuracy and processing efficiency at the same time. It stands out for its speed, surpassing many advanced models, and delivers high accuracy, with a mean average precision (mAP) that exceeds several alternatives. Additionally, its improved design incorporates a new backbone network and an anchor-free detection head, enhancing its segmentation capabilities (12).

The computational complexity of YOLOv12 is self-attention computationally expensive because every element in the input interacts with every other element, making its complexity grow quadratically (L^2). In contrast, CNNs scale more efficiently, processing data linearly (L) using smaller, localized operations. Many attention-based models, like Swin Transformer, add extra complexity, further slowing them down. Additionally, self-attention suffers from memory inefficiencies, as key data frequently moves between high-speed GPU memory (SRAM) and slower main memory (HBM), causing delays. Since SRAM is over 10 times faster than HBM, this memory transfer significantly increases processing time. To address these issues, this paper optimizes attention mechanisms for greater efficiency (42).

2.6.1. YOLOv12 Architectural

YOLOv12 introduces several architectural enhancements to improve computational efficiency and accuracy. Unlike plain-style vision transformers, it retains the hierarchical design of previous YOLO systems. The backbone is simplified by replacing the stacked three-block structure with a single R-ELAN block, optimizing performance. The first two stages are inherited from YOLOv11, without incorporating the newly proposed R-ELAN (42).

To refine the attention mechanism, YOLOv12 adjusts the MLP ratio (from 4 to 1.2 or 2 for smaller models), replaces nn.Linear + LN with nn.Conv2d + BN for better convolution efficiency, and removes positional encoding. Additionally, a 7×7 separable convolution (position perceiver) is introduced to enhance positional awareness. These modifications improve both speed and accuracy, making YOLOv12 more efficient for real-time detection tasks (42).

2.6.2. Hyperparameter Settings

The YOLOv12 model was trained for 140 epochs using a batch size of 32 × 8 to balance GPU memory efficiency with stable gradient updates. An initial learning rate of 10^{-2} was applied and gradually reduced to 10^{-4} using a linear decay scheduler to optimize learning progression. Stochastic Gradient Descent (SGD) was used as the optimizer, and cross-entropy loss served as the loss function, with the learning rate decaying by a factor of 0.1 every seven epochs. The same training configuration was used for the transfer learning model to maintain consistency across evaluations. All hyperparameters were standardized to ensure a fair comparison between the object detection and transfer learning approaches.

2.7. Performance Metrics

To assess the effectiveness of the segmentation techniques performed, four measurements have been used, i.e., Precision, Sensitivity, Specificity, Accuracy, and mean average precision (mAP). That performance metrics can be defined by TP, FP, and FN. the formulas for the main evaluation metrics:

1. Accuracy: Measures the overall correctness of the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision: Measures how many of the predicted positive instances are actually positive.

$$\text{Precision} = \frac{TP}{TP + FP}$$

3. Recall: Measures how many actual positive instances were correctly predicted.

$$\text{Recall} = \frac{TP}{TP + FN}$$

4. F1-Score: The harmonic mean of precision and recall, balancing both.

$$\text{F1} - \text{Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. Mean Average Precision (mAP): a way to evaluate how an object detection model works.

$$mAP = \frac{1}{n}\sum_{k=1}^{n} AP_k$$

Where AP_k is the AP of class k, and n is the number of classes used.

Mean Average Precision (mAP) It’s a method used to assess how well an object detection model performs.It does this by comparing the boxes of the predicted model to the object location in the ground truth image. A prediction is considered accurate if it closely matches the actual object’s location, which is measured using a metric called Intersection over Union (27).

To calculate mAP, the precision and recall of the model are analyzed at different IoU thresholds, and a curve is created. The Average Precision (AP) is later determined for each object class by measuring the area under this curve. Finally, mAP is found by averaging the AP scores for all detected classes. The mAP score ranges from 0 to 1, where higher values mean better accuracy. A typical model would achieve a score of 1.0, while a model that misses all objects would score 0.0.

3. Results and Discussion

To evaluate the effectiveness of the two applied segmentation methods - SS-KM and YOLOv12 -a comprehensive set of performance metrics was used, including accuracy, precision, recall, mean average precision (mAP), and F1-score. The results highlight notable differences between the traditional and deep learning-based approaches, with the YOLOv12 model demonstrating superior performance in terms of detection accuracy and consistency across various cases of dental caries. The following subsections clearify these results.

3.1. YOLOv12 Results (Case1 and case2)

The YOLOv12 model was evaluated in two different cases, each with varying training and testing data splits, to understand the impact of data distribution on its performance, as shown in **Table 3.**

Table 3. Data Splits for Training, Testing, and Validation for YOLOv12 Models.

Method	**Evaluating machine learning models**			
	Training	**Testing**	**Validation**	**Accuracy**
YOLOv12 - Case 1	74%	21%	5%	98.12%
YOLOv12 - Case 2	64%	32%	5%	96.79%

- YOLOv12 - Case 1: In this case, 74% of the data was used for training, 21% for testing, and 5% for validation. The model achieved an accuracy of 96.79%, with precision at 96.7%, recall at 98.7%, and a mean average precision (mAP) of 99.5%. The F1-score for this configuration was 0.99, indicating a near-perfect balance between precision and recall, as shown in **Table (4).**
- YOLOv12 - Case 2: For the second case, the training data was reduced to 64%, with 32% used for testing and 5% for validation. Despite a smaller training set, the model improved its overall accuracy to 98.12%. Precision increased to 99.6%, while recall was slightly lower at 98.1%. The model maintained a high mAP of 99.5% and an F1-score of 0.99, indicating that even with less training data, the model retained excellent performance, **Table (4**) explains the Performance metrics, and figure 3 shows the mAP of both cases, shown in.

These results suggest that YOLOv12 is highly robust to changes in the data split and continues to perform exceptionally well across different configurations, with Case 2 yielding the best overall performance. **Figure (2)** demonstrates this results.

Table 4. Performance metrics for YOLOv12 and MSS-KM Approaches

Method	**metrics**				
	Accuracy	**Precision**	**Recall**	**F1-score**	**mAP**
YOLOv12 - Case 1	**98.12%**	**99.6%**	98.1%	**99.5%**	**0.99**
YOLOv12 – Case 2	96.79%	96.7%	**98.7%**	99.3%	0.99
MSS-KM approach	89.7%	88.5%	91%	89.1%	-

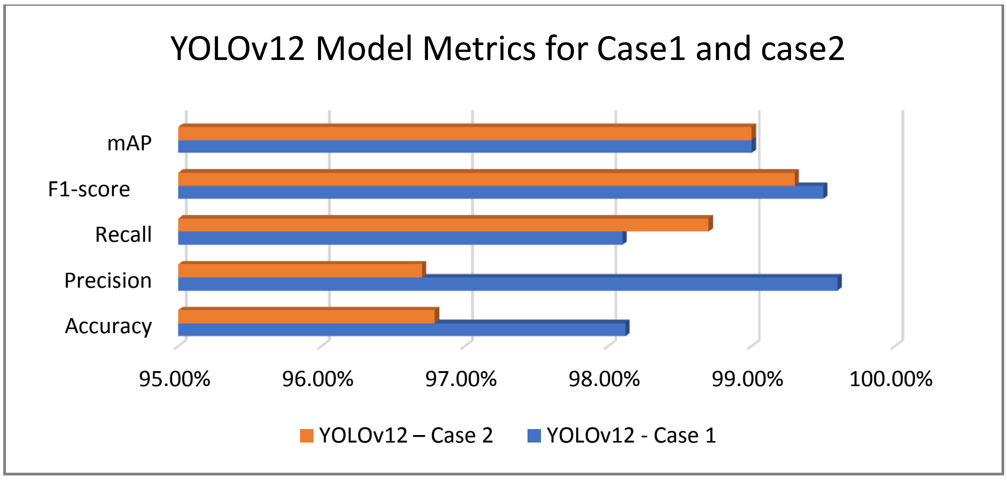

Figure 2. Metrics for YOLOv12 - case1, and YOLOv12 – case2.

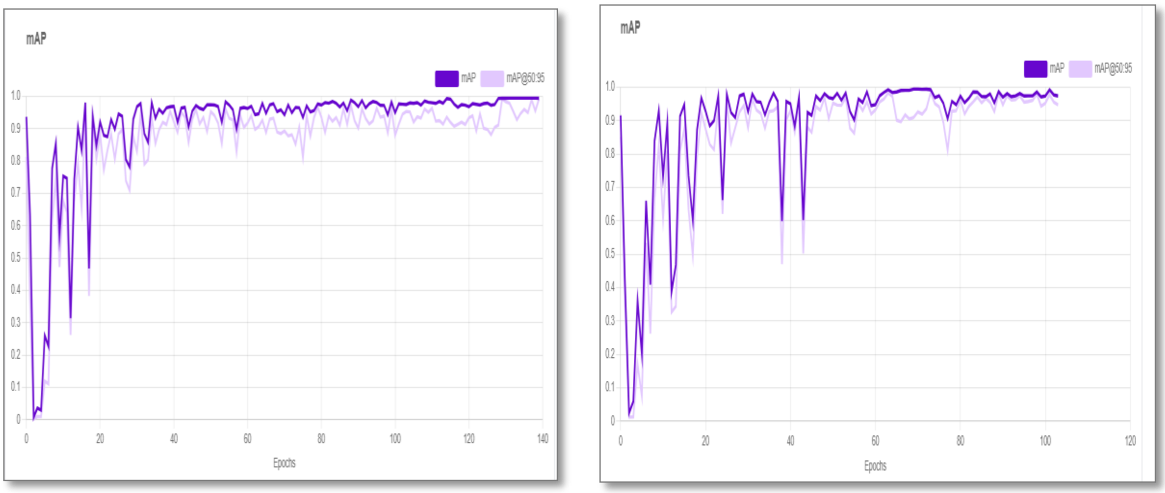

Figure 3. mAP metric for YOLOv12 model, case1, and case 2 respectively.

3.2. MSS-KM Approach Results

On the other hand, the MSS-KM approach, which relies on traditional segmentation techniques, performed at a lower level, as shown in table 4. It achieved an accuracy of 89.7%, precision of 88.5%, and recall of 91%. Its F1-score was 89.1%. However, mAP couldn't be calculated for this method, so we can't compare it directly with YOLOv12 on that metric. While MSS-KM performed reasonably well, it didn't quite reach the performance of YOLOv12. Figure 4 depicts Metrics Parameters for MSS-KM approach.

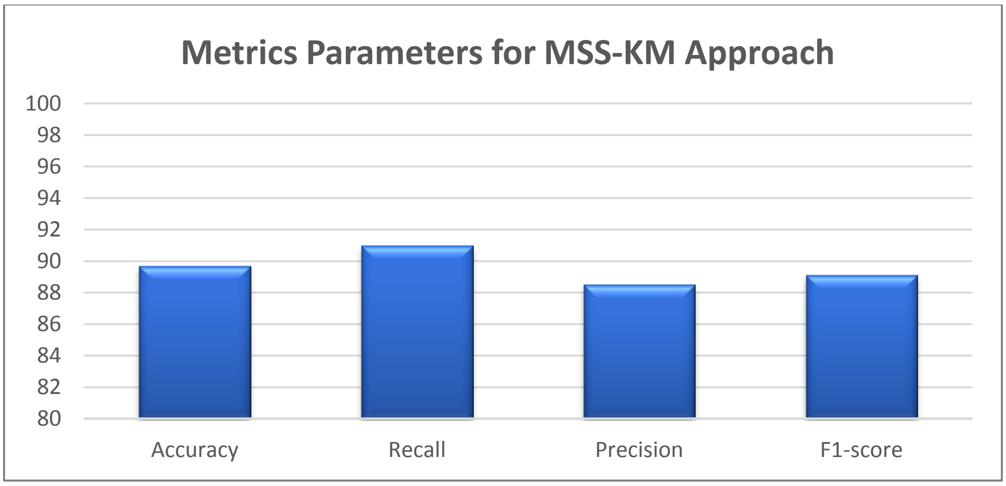

Figure 4. Metrics Parameters MSS-KM approach.

3.3. Analysis and Discussion

The evaluation of both approaches; YOLOv12 and MSS-KM, revealed clear differences in performance. In Table 2, the results from two YOLOv12 training cases are compared against the MSS-KM method. YOLOv12 outperformed the traditional MSS-KM approach across all measured metrics.

In Case 1 of YOLOv12, which used a larger portion of data for training (74%), the model achieved an accuracy of 96.79%, with a precision of 96.7%, recall of 98.7%, and a mean average precision (mAP) of 0.99. In Case 2, where less training data (64%) was used but more was allocated for testing (32%), the model still performed slightly better, achieving 98.12% accuracy, 99.6% precision, 98.1% recall, and the same mAP of 0.99. The F1-scores in both cases were nearly perfect, indicating strong overall performance in detecting dental caries.By contrast, the MSS-KM approach resulted in 89.7% accuracy, 88.5% precision, 91% recall, and an F1-score of 89.1%. Although this method still provided acceptable results, it clearly lagged behind the deep learning-based YOLOv12 model in terms of detection accuracy and consistency.

A key factor contributing to YOLOv12's superior performance lies in its training configuration. The model was trained using a stochastic gradient descent (SGD) optimizer, with a learning rate that started at 10^{-2} and gradually reduced to 10^{-4} using a linear decay strategy. This helped improve convergence during training. A batch size of 32×8 was chosen to balance the GPU memory load while maintaining stable updates to the model's weights. These training settings, aligned with best practices in object detection tasks, were essential in achieving the high accuracy and precision observed in the results.

Overall, these findings highlight the effectiveness of YOLOv12 for caries detection, especially when paired with the right training setup and data distribution. The results also suggest that deep learning-based methods offer a more reliable solution than traditional segmentation techniques like MSS-KM in medical imaging tasks.

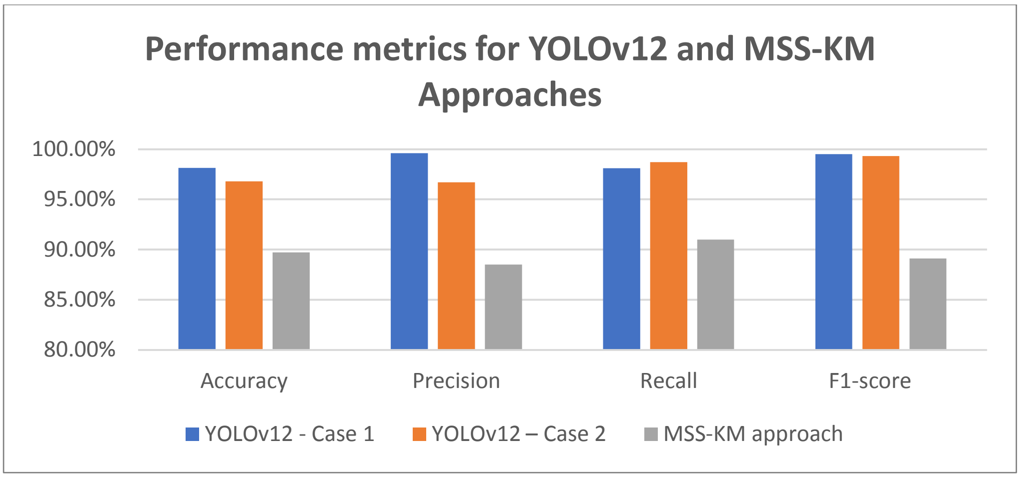

Figure 5. Performance metrics for YOLOv12 and MSS-KM Approaches.

To evaluate the effectiveness of the proposed system, its accuracy was compared with several existing methods from recent literature. **Figure (6)** summarizes the comparison based on reported accuracy values. As shown, the proposed system achieved an accuracy of 98.12%, which is competitive with and in some cases surpasses existing approaches.

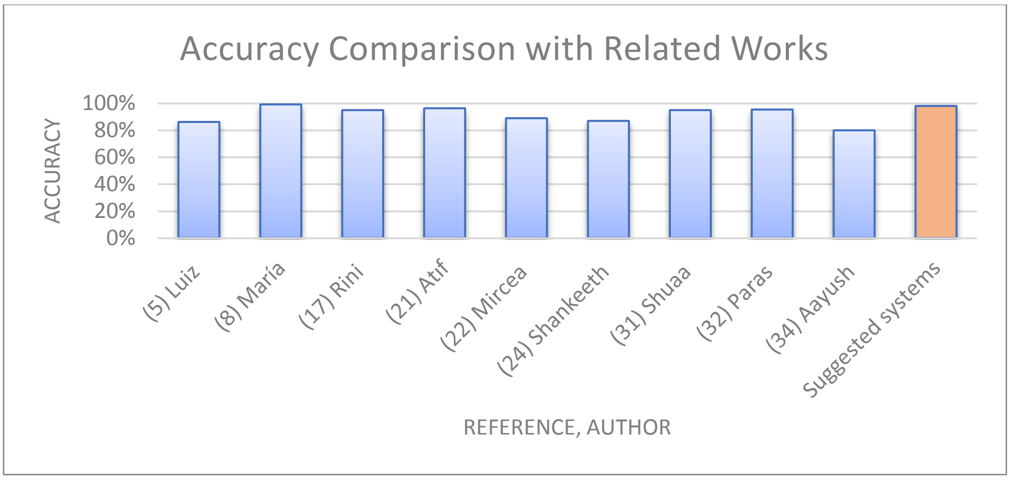

Figure 6. Accuracy comparison with related works.

The results indicate that the proposed method offers high accuracy, placing it among the top-performing systems. While some methods e.g., reference (8), (30) reported slightly higher accuracy, variations in datasets and evaluation protocols must be considered when interpreting these outcomes.

4. Conclusion

This study demonstrated that YOLOv12 significantly outperforms the MSS-KM approach in detecting dental caries from color images. The deep learning model achieved higher accuracy, precision, recall, F1-score, and mAP, referring its effectiveness and reliability for caries detection tasks. The model's ability to produce consistent results across different experimental setups highlights its robustness and adaptability. These outcomes underscore the potential of advanced deep learning models to support more accurate and efficient dental diagnostics. In the future work, one important direction is to expand the dataset to include more diverse and higher-resolution dental images, which would help improve the model's generalization. Additionally, use other types of dental imaging such as X-rays or 3D scans to provide complementary information, especially for detecting early stage caries that are not clearly visible in color images. Finally, integrate other some models to improve model's performance.

Acknowledgment

Our researcher extends his Sincere thanks to the editor and members of the Ibn AL-Haitham Journal of Pure and Applied Sciences preparatory committee.

Conflict of Interest

The authors declare that they have no conflicts of interest.

Funding

This work is self-funding.

References

1. WHO highlights oral health neglect affecting nearly half of the world's population. 2022. Available at: https://www.who.int/news/item/18-11-2022-who-highlights-oral-health-neglect-affecting-nearly-half-of-the-world-s-population
2. Peres M.A., Macpherson L.M.D., Weyant R.J., et al. Oral diseases: A global public health challenge. Lancet 2019; 394(10194):249–260. https://doi.org/10.1016/S0140-6736(19)31146-8

3. Kaguru G., Ayah R., Mutave R., Mugambi C. Integrating oral health into primary health care: A systematic review of oral health training in sub-Saharan Africa. J. Multidiscip. Healthc. 2022; 15:1361–1367. https://doi.org/10.2147/JMDH.S357863
4. Žibienė M., Gaigalienė R., Vaičiulienė I., Skučienė K., Rima J., Saldūnaitė E. Prevalence and severity of dental caries among 18-year-old Lithuanian adolescents. Medicina 2016; 52:54–60
5. Zanini L., Rubira-Bullen I., Nunes F. Segmentation and classification of dental caries in cone beam tomography images using machine learning and image processing. In: Proc. Int. Joint Conf. Biomed. Eng. Syst. Technol. 2024; 2:428–435. https://doi.org/10.5220/0012365300003657
6. Tafere Y., Chanie S., Dessie T., Gedamu H. Assessment of prevalence of dental caries and associated factors among patients attending dental clinic in Debre Tabor general hospital. BMC Oral Health 2018; 18(1):119. https://doi.org/10.1186/s12903-018-0581-8
7. Sonis S.T. Dental Secrets, 4th ed., Elsevier, 2015:147–200. https://doi.org/10.1016/B978-0-323-26278-1.00008-8
8. Prados-Privado M., García Villalón J., Blázquez Torres A., Martínez-Martínez C.H., Ivorra C. A validation employing convolutional neural network for the radiographic detection of absence or presence of teeth. J. Clin. Med. 2021; 10:1186. https://doi.org/10.3390/jcm10061186
9. Thakur G.K., Thakur A., Kulkarni S., Khan N., Khan S. Deep learning approaches for medical image analysis and diagnosis. Cureus 2024; 16(5):e59507. https://doi.org/10.7759/cureus.59507
10. Liu L., Cheng J., Quan Q., Wu F.X., Wang Y.P., Wang J. A survey on U-shaped networks in medical image segmentations. Neurocomputing 2020; 409:244–258
11. Zuo Q., Chen S., Wang Z. R2AU-Net: Attention recurrent residual convolutional neural network for multimodal medical image segmentation. Secur. Commun. Netw. 2021. https://doi.org/10.1155/2021/6625688
12. Bayati M., Alizadeh Savareh B., Ahmadinejad H., et al. Advanced AI-driven detection of interproximal caries in bitewing radiographs using YOLOv8. Sci. Rep. 2025; 15:4641. https://doi.org/10.1038/s41598-024-84737-x
13. Di S., Liao M., Zhao Y., Li Y., Zeng Y. Image superpixel segmentation based on hierarchical multi-level LI-SLIC. Opt. Laser Technol. 2021; 135:106703. https://doi.org/10.1016/j.optlastec.2020.106703
14. Shoaib A., Vadiveloo M., Lim S.P. Comparative studies of region-based segmentation algorithms on natural and remote sensing images. ITM Web Conf. 2024; 67. https://doi.org/10.1051/itmconf/20246701048
15. Soylu E., Guzel M.S., Bostanci G.E., Ekinci F., Asuroglu T., Acici K. Deep-learning-based approaches for semantic segmentation of natural scene images: A review. Electron. 2023; 12:2730. https://doi.org/10.3390/electronics12122730.
16. Archana R., Eliahim Jeevaraj P.S. Deep learning models for digital image processing: A review. *Artif. Intell. Rev.* 2024; 57(11). https://doi.org/10.1007/s10462-023-10631-z
17. Widyaningrum R., Candradewi I., Aji N.R., Aulianisa R. Comparison of multi-label U-Net and Mask R-CNN for panoramic radiograph segmentation to detect periodontitis. *Imaging Sci. Dent.* 2022; 52(4):383–391. https://doi.org/10.5624/isd.20220105
18. Ibraheem N.A., Hasan M.M. Combining several substitution cipher algorithms using circular queue data structure. *Baghdad Sci. J.* 2020; 17(4):1320–1327. https://doi.org/10.21123/bsj.2020.17.4.1320
19. Wiedermann W., Li X. Direction dependence analysis: A framework to test the direction of effects in linear models with an implementation in SPSS. *Behav. Res.* 2018; 50:1581–1601. https://doi.org/10.3758/s13428-018-1031-x
20. Sivari E., Senirkentli G.B., Bostanci E., Guzel M.S., Acici K., Asuroglu T. Deep learning in diagnostic of dental anomalies and diseases: A systematic review. *Diagnostics* 2023; 13:2512. https://doi.org/10.3390/diagnostics13152512
21. Yüksel A.E., Gültekin S., Simsar E., et al. Dental enumeration and multiple treatment detection on panoramic X-rays using deep learning. *Sci. Rep.* 2021; 11:12342. https://doi.org/10.1038/s41598-021-90386-1

22. Muresan M.P., Barbura A.R., Nedevschi S. Teeth detection and dental problem classification in panoramic X-ray images using deep learning and image processing techniques. In: *Proc. IEEE Int. Conf. Intell. Comput. Commun. Process. (ICCP)*, 2020:457–463. https://doi.org/10.1109/ICCP51029.2020.9266244
23. Vasdev D., Gupta V., Shubham S., et al. Retracted article: Periapical dental X-ray image classification using deep neural networks. *Ann. Oper. Res.* 2023; 326(Suppl 1):161. https://doi.org/10.1007/s10479-022-04961-4
24. Vinayahalingam S., Kempers S., Limon L., et al. The automatic detection of caries in third molars on panoramic radiographs using deep learning: A pilot study. *Preprint* 2021. https://doi.org/10.21203/rs.3.rs-379636/v1
25. Khalaf M., Dhannoon B.N. MSRD-Unet: Multiscale residual dilated U-Net for medical image segmentation. *Baghdad Sci. J.* 2022; 19(6):1603–1611. https://doi.org/10.21123/bsj.2022.7559
26. Kim S., Lee C.-O. Individual tooth segmentation in human teeth images using pseudo edge-region obtained by deep neural networks. *Signal Process. Image Commun.* 2024; 120:117076. https://www.sciencedirect.com/science/article/pii/S0923596523001583
27. Tareq A., Faisal M.I., Islam M.S., et al. Visual diagnostics of dental caries through deep learning of non-standardised photographs using a hybrid YOLO ensemble and transfer learning model. *Int. J. Environ. Res. Public Health* 2023; 20(7):5351. https://doi.org/10.3390/ijerph20075351
28. Beser B., Reis T., Berber M.N., et al. YOLO-V5 based deep learning approach for tooth detection and segmentation on pediatric panoramic radiographs in mixed dentition. *BMC Med. Imaging* 2024; 24:172. https://doi.org/10.1186/s12880-024-01338-w
29. Majanga V., Viriri S. A survey of dental caries segmentation and detection techniques. *Sci. World J.* 2022; 5:19. https://doi.org/10.1155/2022/8415705
30. Chawla R., Krishna K.H., Deshmukh A.A., Sagar K.V.D., Al Ansari M.S., Taloba A.I. A hybrid optimization approach with deep learning technique for the classification of dental caries. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* 2022; 13(12).
31. Alharbi S.S., AlRugaibah A.A., Alhasson H.F., Khan R.U. 2023. Detection of cavities from dental panoramic X-ray images using nested U-Net models. *Appl. Sci.* 13:12771. https://doi.org/10.3390/app132312771
32. Tripathi P., Malathy C., Prabhakaran M. 2019. Genetic algorithms based approach for dental caries detection using back propagation neural network. *Int. J. Recent Technol. Eng.* 8(1S2). https://www.ijrte.org/portfolio-item/a00730581s219
33. Jusman Y., Anam M.K., Puspita S., Saleh E. 2021. Machine learnings of dental caries images based on Hu moment invariants features. *Proc. Int. Sem. Technol. Inf. Commun. (iSemantic)*:296–299. https://doi.org/10.1109/iSemantic52711.2021.9573208
34. Aayush J., Himanshu T., Ujjwal K.S., Nitin K., Sandeep K. 2023. Dental caries detection using Faster R-CNN and YOLO V3. *ITM Web Conf.* 53:02005. https://doi.org/10.1051/itmconf/20235302005
35. Zhang S., Ma Z., Zhang G., Lei T., Zhang R., Cui Y. 2020. Semantic image segmentation with deep convolutional neural networks and Quick Shift. *Symmetry* 12(3):427. https://doi.org/10.3390/sym12030427
36. Li X.X., Shen X.J., Chen H.P., et al. 2017. Image clustering segmentation based on SLIC superpixel and transfer learning. *Pattern Recognit. Image Anal.* 27:838–845. https://doi.org/10.1134/S1054661817040101
37. Faragallah O.S., El-Hoseny H.M., El-Sayed H.S. 2023. Efficient COVID-19 super pixel segmentation algorithm using MCFO-based SLIC. *J. Ambient Intell. Humaniz. Comput.* 14(7):9217–9232. https://doi.org/10.1007/s12652-022-04425-4
38. Sasmal B., Dhal K.G. 2023. A survey on the utilization of superpixel image for clustering based image segmentation. *Multimed. Tools Appl.* https://doi.org/10.1007/s11042-023-14861-9
39. Ghosh P., Mali K., Das S.K. 2018. Use of spectral clustering combined with normalized cuts (N-Cuts) in an iterative k-means clustering framework (NKSC) for superpixel segmentation with

contour adherence. *Pattern Recognit. Image Anal.* 28(3):400–409. https://doi.org/10.1134/S1054661818030161

40. Qiao N., Di L. 2022. An improved method of linear spectral clustering. *Multimed. Tools Appl.* 81:1287–1311. https://doi.org/10.1007/s11042-021-11459-x
41. Siyuan R., Xinying L. 2020. Superpixel image segmentation based on improved K-means. *J. Phys. Conf. Ser.* 1533:032067
42. Tian Y., Ye Q., Doermann D. 2025. YOLOv12: Attention-centric real-time object detectors. *arXiv* preprint arXiv:2502.12524. https://doi.org/10.48550/arXiv.2502.12524
43. Osman F.M., Yap M.H. 2020. Adjusted Quick Shift phase preserving dynamic range compression method for breast lesions segmentation. *Inform. Med. Unlocked* 20:100344. https://doi.org/10.1016/j.imu.2020.100344
44. Mashhadani S. 2025. Detection the topics of Facebook posts using text mining with Latent Dirichlet Allocation (LDA) algorithm. *Ibn AL-Haitham J. Pure Appl. Sci.* 38(1). https://doi.org/10.30526/38.1.4033
45. Abdalah A.M., Alkazaz N.R. 2024. Predicting movie production years through facial recognition of actors with machine learning. *Baghdad Sci. J.* 21(12):Art.30. https://doi.org/10.21123/bsj.2024.8996
46. Kaggle. Oral diseases dataset. https://www.kaggle.com/datasets/salmansajid05/oral-diseases