Science Journal of University of Zakho
DOI: https://doi.org/10.25271/sjuoz.2026.14.3.1713

Original Article

ENHANCED DENTAL DISEASE DIAGNOSIS USING YOLO11 AND AN INTEGRATED DATASET

Mayav A. Abdulmajeed [1,*], and Mohammed A. Shakir[1]

[1]Faculty of Engineering, University of Duhok-Duhok, Kurdistan Region, Iraq.

[*]*Corresponding author, E-mail:*mayav.aa4822@stu.uod.ac (Tel: +964-7503444355)

ABSTRACT

Received
July 17, 2025

Accepted
August 28, 2025

Published
July 3, 2026

Detection of dental diseases is a crucial initial step in dental treatment's planning. The most common dental diseases are cavity, filling, impacted and implants. Accurate diagnosis is the most important aspect of dental diseases' treatment. Various deep learning (DL) models are used for the detection of dental diseases. Panoramic radiographic images are used widely for detecting dental cavity. The overall performance of the diagnostic models that use panoramic radiographic images for the cavity detection remains significantly challenging. This paper utilizes the Dental Radiography Dataset (DRD), the Real Image Dataset (RID), and a proposed-preprocessed- integrated dataset (the DRD is merged with the RID) in YOLO11's DL model. When YOLO11 trained on the DRD, it provided detection enhancement for all classes. However, the cavity detection remained low. The cavity class was then removed from the DRD, and the RID was replaced to overcome the problem of low detection. The DRD consisted of 1269 images. The RID consisted of 3340 images after augmentation (flip, rotation, and zoom). The mean average precision (mAP) when the intersection over union (IOU) is greater than or equal to 0.5 (mAP@50) is determined to evaluate the performance of the model. The YOLO11 with DRD achieved mAP@50 of 83.7%, and after removing the cavity class from DRD, mAP@50 reached 93.6%. The YOLO11, when trained with the RID, obtained mAP@50 of 85.8%, and post-augmentation, it increased to 89.9%. The integrated dataset produced mAP@50 of 90.4%. The preprocessed integrated dataset demonstrated superior detection performance compared to results in previous studies.

Keywords: Panoramic radiographs; Radiographic images; Integrated dataset; Deep learning; Object detection; Real image dataset; Dental diseases; YOLO.

1. INTRODUCTION

The teeth are rigid calcified structures located in the jaws. Several infections deteriorate dental health such as impacted teeth, dental cavities, root decay, abscessed teeth, and bone loss. Cavity is recognized as the most widespread non-communicable disease, causing permanent damage to tooth structures that can ultimately lead to the formation of holes. Filling is the method for restoring damaged tooth structures (George *et al.*, 2023). Caries may exacerbate periodontal disease (gum disease) and increase the risk of developing cardiovascular disease which could be connected to other health issues. An essential component of caries therapy is always early diagnosis (Alkentor & Assalem, 2024). Dental implant is a common and effective treatment option for individuals who have lost teeth due to decay, trauma, or other reasons (Heimes *et al.*, 2023). An impacted tooth is a partially erupted or remains unerupted, resting against another tooth, bone, or soft tissue (Watted *et al.*, 2025).

Radiological assessments play an essential role in the diagnosis of dental diseases. Various types of radiographic images are utilized in the detection of dental diseases (Karakuş *et al.*, 2024). Panoramic dental radiographic images are crucial diagnostic and therapeutic tools for patients used by dentists (Beser *et al.*, 2024). The extra-oral radiographic techniques have the advantage of utilizing low-dose radiation to provide rapid and simple radiological imaging (Kurt-Bayrakdar *et al.*, 2024). Panoramic X-ray image contains information about the mouth, such as structures, teeth, lower and upper jaws, and surrounding tissues; the information appears in two-dimensional (2-D) (Sunilkumar *et al.*, 2024).

Dental caries can be challenging to identify with radiographs, despite the fact that they are generally acknowledged as a detection tool for the presence of caries lesions. The factors that may impact the inter-rater agreement include the radiograph resolution, the dentist's perspective, and the duration of each examination (Al-Ghamdi *et al.*, 2022). As a result, all dentists must adhere to the protocol when interpreting results, ascertain whether the patient has dental caries or not; therefore, it is necessary to examine the tooth (Jusman *et al.*, 2021). To address dental caries, a real image dataset can be used by dentists.

Dentists sometimes expect a significant number of patients to treat in a day, resulting too many dental X-ray images daily as a crucial diagnostic tool. Interpreting X-rays consumes time and raises the possibility of incorrect or diagnostic oversight due to personal causes such as fatigue, insufficient skill, and emotions. As a result, the treatment may occasionally be hampered by depending on dentists (Thulaseedharan & PS, 2023). Artificial intelligence (AI) systems have the potential to interpret large amount of data properly and fast, producing insights that would be difficult to obtain by human (Nour *et al.*, 2025). The deep learning (DL) techniques have significantly been improved in dental medicine, particularly in tooth identification problems. However, DL models' effectiveness is largely dependent on the quality and the quantity of training datasets. Image augmentation is a critical

Access this article online

DOI:https://doi.org/10.25271/sjuoz.2026.14.3.1713

Printed ISSN 2663-628X;
Electronic ISSN 2663-6298

Science Journal of University of Zakho
Vol. 14, No. 03, pp. 410-425, July -2026

approach for addressing issues caused by insufficient data. Diverse changes and variations enhance the generalization capability and robustness of deep learning models across different tasks related to computer vision (CV). The most widely used task in the CV is object detection. Object detection plays a crucial role in accurately localizing and classifying dental diseases (Abbas & Mueen, 2025; Vempati, 2024). There are various image detection models, and YOLO (You Only Look Once), is one of the most notable object detection models widely adopted in dental imaging. YOLO has significantly enhanced the speed of detection tasks and its single pipe identification (Özçelik *et al.*, 2025). Various versions of YOLO model have been launched; among them there is YOLO11. YOLO11 has an enhanced architecture that reduces the number of parameters by 22% while achieving impressive performance gains and improved real-time image processing speed. It also enhances feature extraction capabilities and object detection accuracy (Kutyrev *et al.*, 2025).

To address this gap and utilize RID instead of the cavity class in DRD, as the cavity class in panoramic X-ray decreases the performance of the system's overall accuracy. Cavity class possesses better detection in real images than panoramic X-ray images, and this will follow as exact protocol utilized by dentists. The RID can be taken by a mobile phone or any electronic device that has a camera. The RID is easy to be taken, real-time, and free, and does not require an expensive device like panoramic X-rays. Our goal in this paper is to produce a model using DL that can precisely detect common dental infections to help dentists detect early cavities in the most effective way to protect dental health.

This paper employs YOLO11, due to having higher detection with less computational complexity and number of parameters. In the meantime, it replaces the cavity class in DRD with the RID cavity classes, and the new utilized dataset is renamed as the integrated dataset. The integrated dataset, which comprises six classes, i.e., filling, impacted, implant, cavity decay, early decay, and healthy teeth, will be trained on the YOLO11 model.

The rest of this paper is structured as follows. Section 2 presents the literature review, while Section 3 describes the materials and methods. Section 4 outlines the evaluation metrics, and Section 5 reports the experimental results. Section 6 provides a comparative analysis, followed by Section 7, which discusses the findings. Finally, Section 8 concludes the study.

2. LITERATURE REVIEW

A significant amount of research has been conducted on the detection of dental diseases using various DL models across multiple dental image modalities. DL techniques have demonstrated notable improvements in tooth disease detection. Diseases have explored the use of real dental images to enhance the diagnosing accuracy. Real images tend to provide more accurate results than panoramic radiography images for certain classes. Among the advanced object detection models, YOLO has emerged as a widely adopted approach. This review examines current research findings related to tooth disease detection using a panoramic radiographic dataset, an RID, and a hybrid dataset.

Panoramic radiographic dataset

Thulaseedharan and PS (2022) had collected 664 panoramic X-ray images from three clinicians to construct a labeled dataset. They labeled the images as nine categories: crowns, impactions, root stumps, dental caries, implants, restorations, bridges, orthodontic treatment, and root canal treatment. They had applied data augmentation techniques to enhance detection performance; as a result, the YOLOv5 model achieved a mAP@50 of 72.4%. In a subsequent study after one year, they utilized the same dataset, trained a YOLOv6 model for 1,000 epochs, and achieved a mAP@50 of 70.76% (Thulaseedharan & PS, 2023).

George *et al.* (2023) divided their study into three main phases: enhancement, segmentation, and training the YOLOv8 model. They selected a publicly available dataset annotated into four diseases (i.e., cavity, filling, impacted, and implant), containing 1,000 images. After cleaning the data by removing 264 low-quality or irrelevant images, they augmented and split the dataset for training and evaluation. For the enhancement phase, images were processed using upsampling, downsampling, and two additional enhancement techniques. The segmentation process consisted of three stages. Initially, jaw area extraction was applied to remove all regions not containing dental information. After that, upper and lower jaw separation was performed to isolate the upper and lower jaws. Finally, vertical segmentation was used to divide each jaw into two vertical sections, resulting in four distinct regions per image. The YOLOv8 model was trained on the segmented regions. The proposed method achieved a mAP@50 of 82.1%.

Alkentar and Assalem (2024) used two datasets with several models. Dentex and dental radiography datasets (DRD) were utilized which are publicly available. The DRD contained 1075 images annotated with the four diseases: caries, implant, impacted teeth, and filling. Dentex contained four classes (impacted teeth, caries, periapical lesions, and deep caries) with 756 images. They had used three models for training; YOLOv7, YOLOv8, and YOLOv9. The methodology has been divided into three stages. An initial step to apply the YOLO model to teeth images. The second step is cropping each tooth image and applying a filter to it. The final step is applying the residents. The mAP@50 results achieved for the DRD dataset were (40.3, 76.6, and 75.9) % for YOLOv7, YOLOv8, and YOLOv9 respectively.

Maged *et al.* (2024) proposed using YOLOv8 for object detection on a DRD containing 1,267 images. The images were divided into 84.5% for training, 5.7% for testing, and 9.5% for validation. The dataset included four classes: implant, filling, impacted tooth, and cavities. The model was trained for 50 epochs and achieved a mAP@50 of 80.6%.

Real image dataset

Khan *et al.* (2022) collected 609 oral images containing two classes of cavity and healthy teeth from various online websites. The images were resized, converted to grayscale, and then applied to augmentation techniques. A new algorithm called Dental-Net was proposed and applied to the dataset, indicating whether a tooth contains a cavity or not. The proposed model was evaluated against several pre-trained models such as VGG16, MobileNetV2, InceptionV3, and ResNet18. The Dental-Net achieved the highest performance with a validation accuracy of 91.09%.

Tareq *et al.* (2023) used a mobile camera to collect 233 images containing three classes: without cavity, microcavity, and cavity. After cleaning the dataset, 165 images remained and they were split into 65:15:20 for training, validation, and testing, respectively. Thirteen types of augmentation techniques were applied to the dataset. This study employed two types of computer vision tasks: object detection and classification. In object detection, all versions of YOLOv5 were evaluated with and without test-time augmentation (TTA), and YOLOv5m with TTA achieved the highest mAP@50 of 50%. For classification, transfer learning was applied using VGG, densenet121, AlexNet, and resnet50. The VGG achieved the best accuracy of 86.96%.

Combined dataset

Rashid *et al.* (2022) proposed using mixed datasets of panoramic X-rays and colorful dental photographs for diagnostic cavity class only. The colorful images were required to be captured using a specific type of camera. ResNet50 was employed for feature extraction and classification through a transfer learning approach. MaskR-CNN was used as the training model, incorporating a region proposal network to localize potential cavity regions. The precision and recall values on the validation set were 89% and 92% for the colorful image dataset, 79% and 80% for the X-ray dataset, and 76.02% and 78.78% for the mixed dataset, with an IOU threshold of 0.7.

In summary of the literature review, it is observed that the whole researchers employed one of the abovementioned techniques. Firstly, multiple models were applied to various datasets to determine which model outperformed the others. Secondly, different datasets were utilized with the selected models to evaluate their performance comprehensively. Finally, the model performance was optimized through hyperparameter tuning.

3. MATERIALS AND METHODS

In this work, two datasets have been integrated: a Dental Radiography Dataset (DRD), which is a panoramic X-ray dataset produced by Mohamadreza Momeni, stored at (`https://www.kaggle.com/imtkaggleteam`), and a real-image dataset (RID), which was obtained from the Roboflow website (`https://app.roboflow.com`). Both DRD and RID datasets were used as references for training the YOLO11 model. Images in DRD and RID were resized with annotation format conversion for DRD and augmentation for RID (see Figure 1).

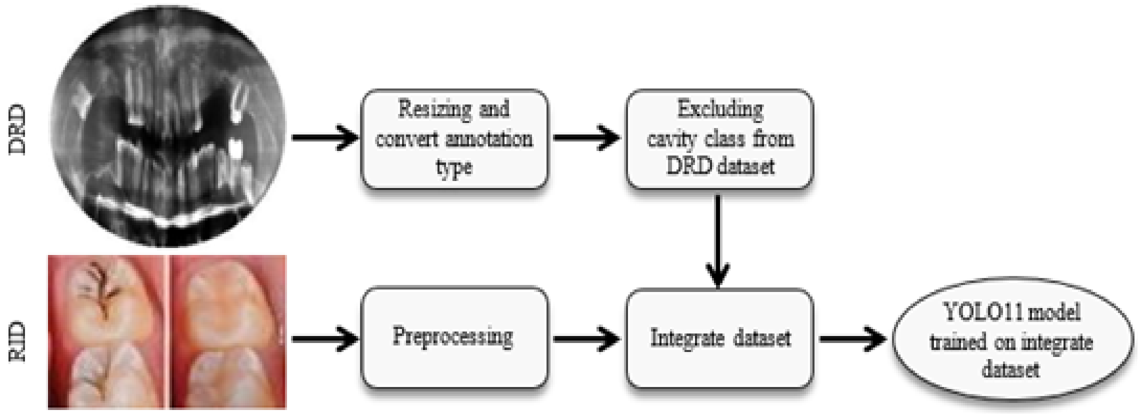


Figure 1: Block diagram illustrating the process of integrating two datasets.

Dental Radiography Dataset (DRD)

The panoramic X-ray image is the most common tool used in the detection of dental diseases. The DRD contains panoramic X-ray images. The DRD contains cavity, filling, impacted, and implant classes of teeth. The dataset is split into three groups: training, validation, and testing. The training contains 1075 original and augmented images. The validation and testing sets contain 121 and 73 images out of the original images. The image size in the DRD is 512*256 pixels. The format of the bounding box annotations is TensorFlow format.

Resizing and annotation format conversion

For the trade between efficiency and real-time detection, the images were resized to 640 × 640 pixels. Bounding box annotations of the DRD images were converted from TensorFlow format (Figure 2) to the YOLO format (Figure 3) to easily train the YOLO11. For the above conversion, the following Equations (1 – 4) were performed:

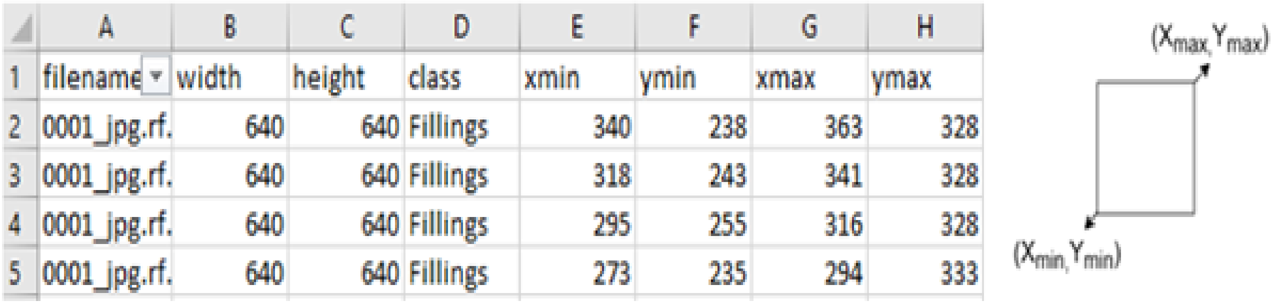
$$Y_{\text{center}} = \frac{Y_{\max} - Y_{\min}}{2} \quad (1)$$

$$X_{\text{center}} = \frac{X_{\max} - X_{\min}}{2} \quad (2)$$

$$\text{Height} = Y_{\max} - Y_{\min} \quad (3)$$

$$\text{Width} = X_{\max} - X_{\min} \quad (4)$$

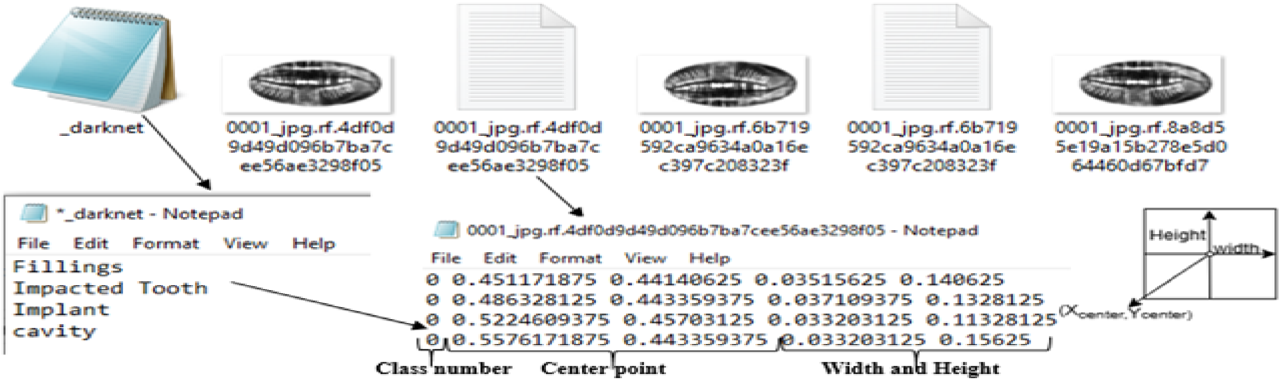
Where: $Y_{\max}$, Y_{center}, and $Y_{\min}$ represent the maximum, center, and minimum point of the height of the bounding box. $X_{\max}$, X_{center}, and $X_{\min}$ represent the maximum, center, and minimum points of the width of the bounding box.



	A	B	C	D	E	F	G	H
1	filename	width	height	class	xmin	ymin	xmax	ymax
2	0001_jpg.rf.	640	640	Fillings	340	238	363	328
3	0001_jpg.rf.	640	640	Fillings	318	243	341	328
4	0001_jpg.rf.	640	640	Fillings	295	255	316	328
5	0001_jpg.rf.	640	640	Fillings	273	235	294	333

Figure 2: Image and bounding box dimensions in TensorFlow format.

After annotation bounding boxes conversion by using the Roboflow website, two types of files are created: a darknet file and a file for each image of the DRD. The darknet file contains class names indicating the dental disease, as shown in Figure 3. Files of the images contain five columns where the first column is linked with Darknet file, such as 0 for fillings, 1 for impacted tooth, 2 for implanted tooth, and 3 for cavity,(Figure 3). The remaining four columns are annotated in Figure 3.



```
Fillings
Impacted Tooth
Implant
cavity
```

```
0 0.451171875 0.44140625 0.03515625 0.140625
0 0.486328125 0.443359375 0.037109375 0.1328125
0 0.5224609375 0.45703125 0.033203125 0.11328125
0 0.5576171875 0.443359375 0.033203125 0.15625
```

Figure 3: Image and bounding box dimensions in TensorFlow format.

Excluding dental cavity images (Cavity-class removal)

During the performance observation of the present data, it was found that detecting dental cavity disease in DRD is difficult due to the low resolution and grey scale image. Therefore, the dental cavity class was removed from the DRD, resulting in a reduction of images in the DRD, as some images contained only cavities; these images with their files were deleted. Hence, the remaining images in the DRD were 1059, 117, and 69 in the training, validation, and testing folders, respectively. For the cavity-class removal, there are two mechanisms: manual removal and software removal. In this work, excluding the dental cavity class from the DRD was achieved by roboflow website.

Real image dataset (RID)

Real image is a state-of-the-art method used to detect the dental health status. RID contains three classes referring to the dental status: healthy, decay cavity, and early decay cavity. RID consists of 1392 original images with a size of 640 × 640 pixels. The dataset is divided into three groups: 988 training, 279 validation and 125 testing images, which are all collected from the Roboflow website. The downloaded dataset consists of raw images that are not augmented.

Preprocessing of the RID

Initially, all images were adjusted to 80*80 pixels. To guarantee the robustness of the evaluation method, the dataset was re-divided into three distinct subsets in a 70:20:10 ratio. As a result, such 974, 279, 139 original images were prepared for training, validation, and testing. To enhance the accuracy and robustness of the model, various data augmentation techniques were applied to the RID, including flipping, brightness, shearing, zooming, rotation, and saturation. These techniques were explored to identify the most effective augmentation strategy for RID. The collection of augmentation strategies are described in Table 4. The highest mAP@50 during training was achieved by applying a combination of horizontal flipping, rotation (90° clockwise, 90° counterclockwise, and 180°), and random zooming (up to 20%) exclusively to the training set. Each 64 images above, resized and augmented images, were combined in one set of 640*640-pixel images as shown in Figure 4. Image size of 640*640 pixels makes a trade between efficiency and real-time detection for training YOLO. Throughout our work, it has been observed that approximately the same results were obtained either with training the individual or the combined real images. However, using the combined set of the RID reduces the training time, produces a high similarity of bounding boxes of the two datasets, detects more than one tooth in one image, and seems to be more similar to the mouth shape than the tooth in the image.

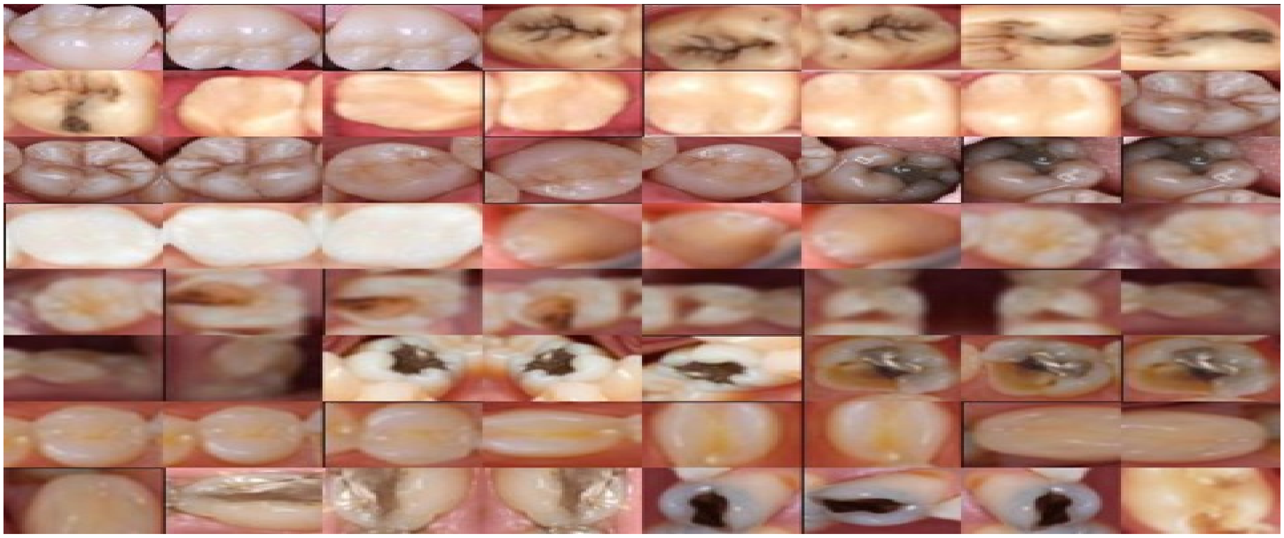


Figure 4: Combined image consisting of 64 images of size 80*80 pixels.

Proposed Method (Integrated Dataset)

For enhancing the detection of all classes, two datasets were integrated and trained with YOLO11 model. The process of integrating two datasets is illustrated in Figure 5. The integrated dataset contains six classes of dental status such as filling, impacted, implant, decay cavity, early decay, and healthy teeth. The two datasets were integrated and processed via the 'Roboflow' website.

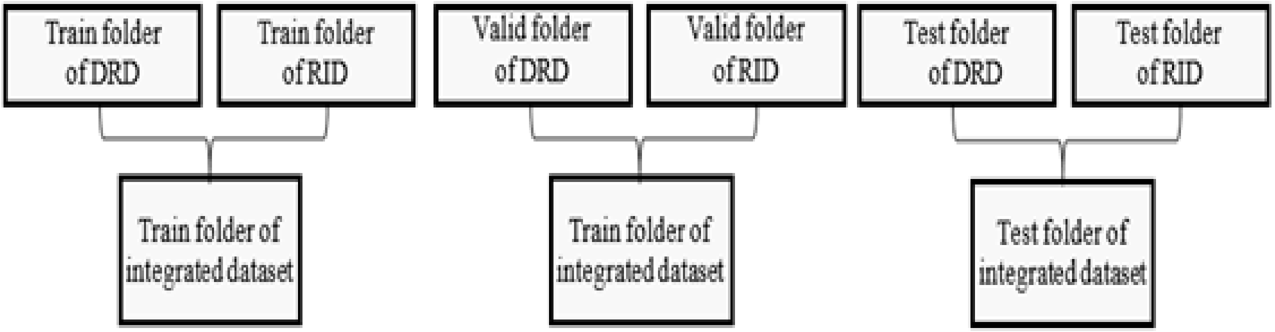
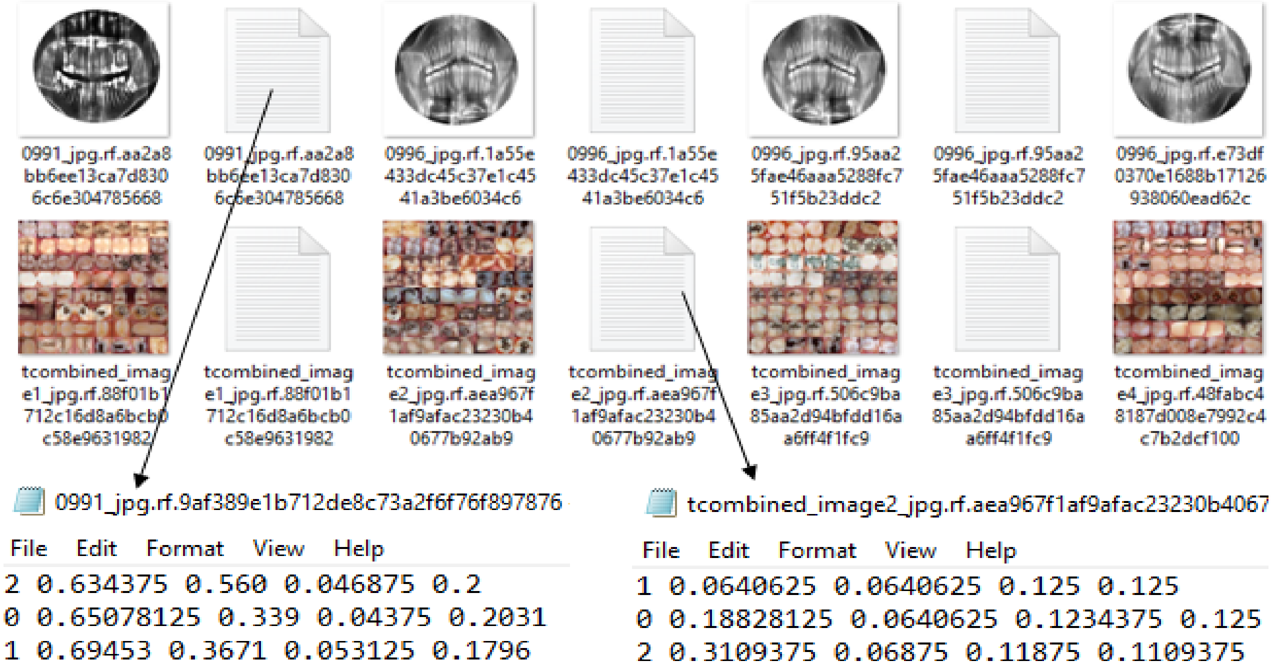


Figure 5: Block diagram of the approach to integrate the two datasets.

Datasets integration mechanism

The idea of the dataset integration (combination) is to copy all images with their annotation files from the training folder of the augmented RID to the DRD training folder. For the validation and testing folder the same procedure is handled. This procedure has two problems: (1) duplication, as darknet file is not allowed, and (2) the overlap between the number of classes of the RID and the DRD, as shown in the Figure 6. This problem can be solved manually or by a software (on website); this problem has been overcome using Roboflow website.



```
2 0.634375 0.560 0.046875 0.2
0 0.65078125 0.339 0.04375 0.2031
1 0.69453 0.3671 0.053125 0.1796
```

```
1 0.0640625 0.0640625 0.125 0.125
0 0.18828125 0.0640625 0.1234375 0.125
2 0.3109375 0.06875 0.11875 0.1109375
```

Figure 6: Idea of the dataset integration.

On the Roboflow website, initially DRD and RID datasets are uploaded separately. Then, the projects have been merged (Figure 7). In the meantime, Roboflow will solve both abovementioned problems automatically (Figure 8).

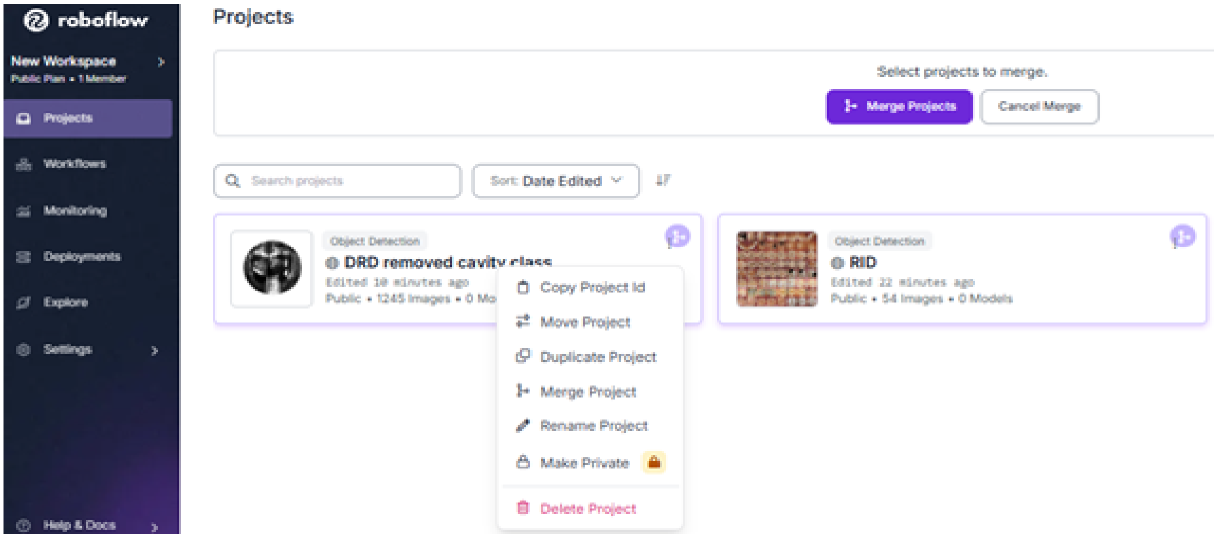
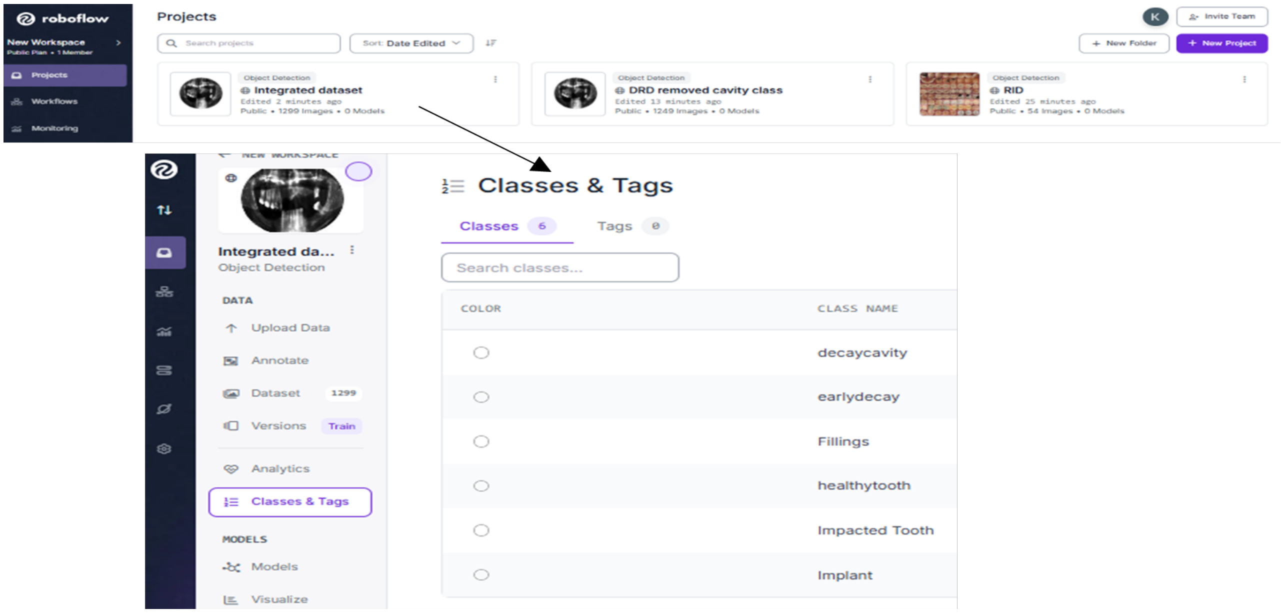


Figure 7: A way of combining two datasets by Roboflow.



COLOR	CLASS NAME
○	decaycavity
○	earlydecay
○	Fillings
○	healthytooth
○	Impacted Tooth
○	Implant

Figure 8: The integrated dataset in roboflow is ready (class and darknet problems handled).

Accurate detection of dental diseases requires different types of tools such as radiography, X-ray, or real images. For instance, implant and impacted teeth detection by real-image is hard, while dental cavity is not precisely detectable via panoramic X-ray. Therefore, dental cavities are relatively easier to detect via real images as the cavity level is based on the spot color and size on the tooth image. Therefore, the dental cavity class on the DRD was excluded and replaced by cavity classes in the RID (early and decay cavities).

In the DRD, low accuracy produces the cavity class in the radiograph images. Additionally, detecting the degree or type of the cavity is too difficult. On the other side, dental cavity detection, with the cavity degree, is more accurate and accessed via RID. Therefore, using two separate systems to accurately diagnose all dental-related diseases is difficult and time-consuming. To overcome this issue, both datasets were integrated to evaluate the overall detection performance.

YOLO11 ARCHITECTURE

The most widely used model for dental disease detection is YOLOv8. The updated or more identical version is YOLO11, which is released on September 30, 2024 by Ultralytics with five versions: n (nano), s (small), m (medium), l (large), and x (x-large). All versions are different in terms of number of parameter and computing complexity (Kumar *et al.*, 2021). YOLO11 has a significant advancement in object detection that combines architectural advances to maximize accuracy and speed detection (Vempati, 2024). YOLO11 architecture contains four parts; the input layer, backbone, neck, and the head. In the input layer, adaptive scaling and data distribution alignment are used to ensure consistency with the training dataset (Ji *et al.*, 2025). The YOLO11 works with square images and uses padded grey pixels to ensure the aspect ratio stability. It will scale without any padding if the image is square. The **backbone** acts as the main feature extractor, converting unprocessed image data into multi-scale feature maps by using several convolutional (Conv) techniques. For example, C3K2.Conv block contains three branches (conv2d, BatchNorm2D, and the SiLU). Their active function acts as a feature extractor. C3K2 is replaced by C2f available in YOLOv8. It is known as the main feature extractor component to extract diversity and multiscale features, and to balance between representing features and computing performance by utilizing residual links and multibranch design. Additionally, it consists of one SPPF (Spatial Pyramid Pooling-Fast) and C2PSA (Parallel Spatial Attention) blocks (Hidayatullah *et al.*, 2025). SPPF is intended to improve object detection capabilities by effectively combining multiscale information. C2PSA block presents YOLOv11's novel attention mechanism which intended to enhance feature representation by pixel- level spatial optimization and cross-scale attention. This block improves the model's ability to identify intricate objects and record minute details. By C3K2 and C2PSA, YOLO11 enhances the feature extraction and detection, and reduces number of parameters. The neck, **which** serves as a processing stage between the backbone and head, uses specialized layers to combine and improve features represented in various scales. The head is the last part, an essential module that generates predictions for detected objects. It accurately predicts the bounding boxes, determines the object class, and computes the objectness score. In head, there are three detection sizes: small, medium, and large. Each one detects objects in a specific size (Rasheed & Zarkoosh, 2025). The block diagram of YOLO11 is presented in Figure 9.

Training setup and hyperparameters

The YOLO11 is trained on the DRD, the DRD with cavity class removed, the RID, the augmented RID, and the preprocessed integrated dataset to find evaluation metrics of each dataset separately. The implementation of YOLO11 model has been run on the Google Colab platform, utilizing Python version 3.11.13. The results were conducted using the Ultralytics YOLO version 8.3.155 framework in conjunction with Torch version 2.6.0+cu124. The utilized parameter was fine-tuned to classify and detect; a perfect hyperparameter configuration consisted of 50 epochs, with a batch size of 16 and AdamW optimization.

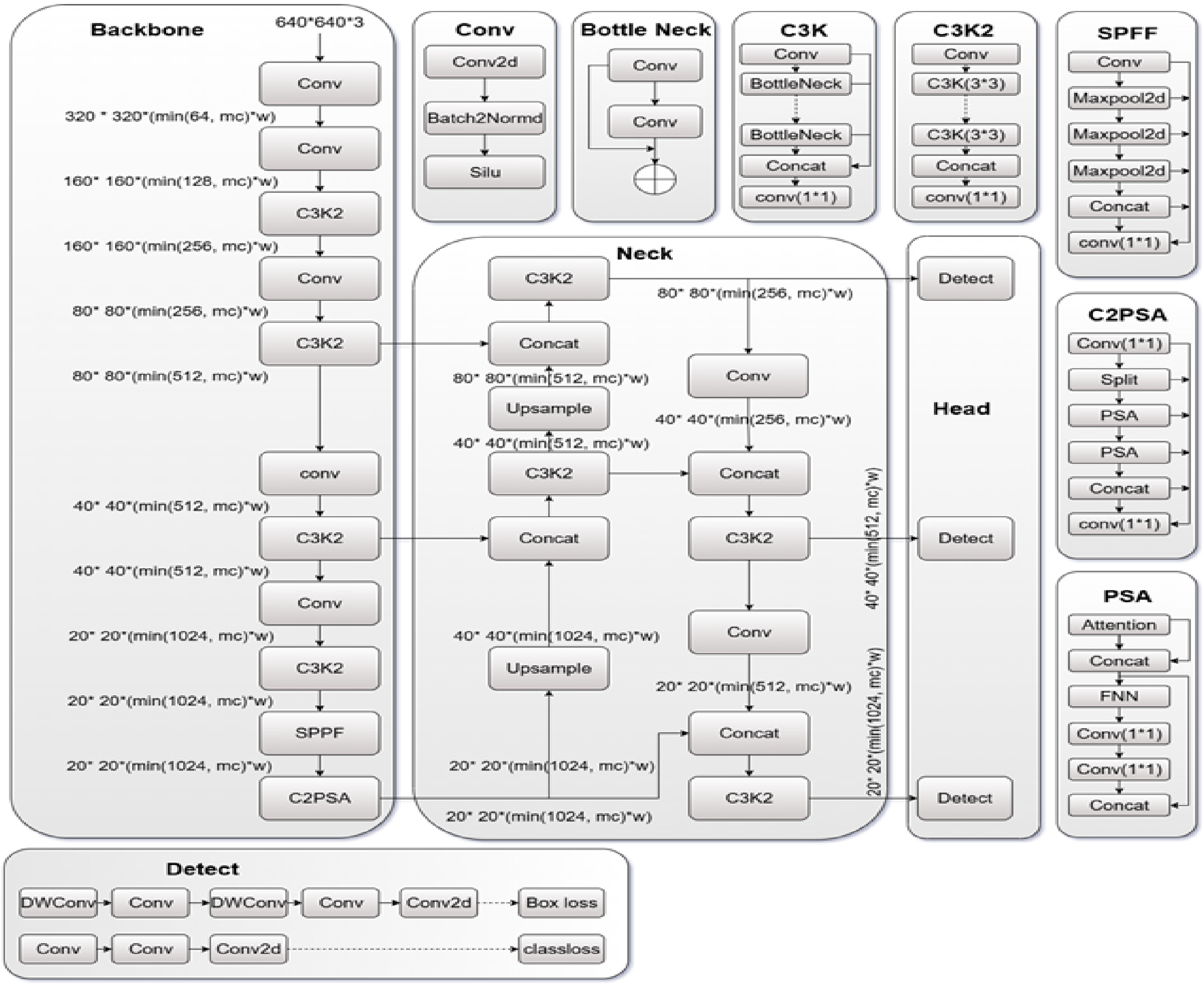


Figure 9: Block diagram of the YOLO11 architecture.

4. EVALUATION METRICS

To evaluate the model performance, three main evaluation metrics are used in this study: precision, recall, and mean average precision (mAP).

The bounding box is extensively employed in dental object detection, which is the most crucial stage for the tooth localization. Two types of bounding boxes are shown in this work: the predicted bounding box that is produced by the YOLO model, and the ground truth bounding box that determines the actual position and dimension of the object and is considered to be the reference for the predicted bounding box. **The intersection over union (IOU)** was performed to measure the overlap between predicted and ground truth bounding boxes, as shown in Figure 10 (Khanam *et al.*, 2025; Padilla *et al.*, 2020). IOU is measured via Equations (5–9).

$$X2 = \min(X2_{\text{predicted}}, X2_{\text{ground truth}}) \tag{5}$$

$$X1 = \max(X1_{\text{predicted}}, X1_{\text{ground truth}}) \tag{6}$$

$$Y2 = \min(Y2_{\text{predicted}}, Y2_{\text{ground truth}}) \tag{7}$$

$$Y1 = \max(Y1_{\text{predicted}}, Y1_{\text{ground truth}}) \tag{8}$$

$$\text{IOU} = \frac{\text{area of intersection}}{\text{area of union}} = \frac{A \cap B}{A \cup B} = \frac{(X2 - X1) * (Y2 - Y1)}{areaA + areaB - A \cap B} \tag{9}$$

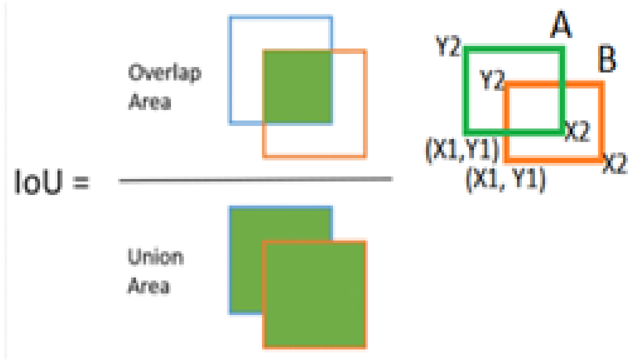


Figure 10: Bounding boxes overlap in resemblance.

Where: X1 is the starting point on the X-axis of the predicted and ground truth bounding boxes. Y1 is the starting point on the Y-axis of the predicted and ground truth bounding boxes. X2 is the finishing point on the X-axis of the predicted and ground truth bounding boxes. Y2 is the finishing point on the Y-axis of the predicted and ground truth bounding boxes.

The method for calculating True positive (TP), False positive (FP), and False negative (FN) is as follows: the result will be considered as a TP if the IOU is greater than or equal to 0.5 which classifies the actual class of dental disease. The outcome is determined as FP if the above value is less than 0.5 or the predicted class does not match the actual class. Finally, if the model (derived from the YOLO) could not detect the dental disease, it is allocated to be (FN).

Precision (P) is the proportion of correctly predicted positive observations to all positive observations, according to Equation (10).

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10}$$

noindent Recall is calculated by dividing the total number of observations by the number of correctly predicted positive outcomes (Logabiraman *et al.*, 2024). Recall can be measured using Equation (11).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{11}$$

Based on the above equations, average precision (AP) is calculated by the plotting and integration the result of Equations (10 and 11) in order to measure the comprehensive assessment metric that will impartially assess object identification models' performance across various IOU criteria as demonstrated in Equation (12):

$$\text{AP} = \int_0^1 p(r)\,dr \tag{12}$$

Mean average precision (mAP) offers a comprehensive measure for assessing object detection performance across several classes and IOU levels. A mAP, which is derived by averaging the AP for each class, presents the balance between precision and recall across various confidence thresholds (Zaidi *et al.*, 2022). It is calculated by applying Equation (13).

$$mAP = \left(\frac{1}{N}\right) * \sum_{i=1}^{N} (\text{AP}_i) \tag{13}$$

The harmonic mean of precision and recall (F1) score is a widely used metric for assessing object detection models. It takes into account both recall and precision of the model to determine a score ranging from 0 to 1 (Li *et al.*, 2023). It is calculated by applying Equation (14).

$$\text{F1} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{14}$$

Box loss (Box_{loss}) plays an important role in the localization accuracy and directly impacts it. It determines the discrepancy between the expected bounding box and the ground truth (Behzadi *et al.*, 2025). It is calculated by applying Equation (15 – 17).

$$Box_{\text{loss}} = 1 - CIOU = 1 - \left[IOU - \frac{\rho^2(b, b^*)}{c^2} - \alpha\, v \right] \tag{15}$$

$$v = \frac{4}{\pi^2} (\arctan \frac{w^*}{h^2} - \arctan \frac{w}{h})^2 \tag{16}$$

$$\alpha = \frac{v}{(1 - IOU) + v} \tag{17}$$

Classification loss (Cls_{Loss}) is a crucial statistic for maximizing the model's object classification performance. It uses weighted binary cross-entropy (WBCE) loss (Behzadi *et al.*, 2025). It is calculated by applying Equation (18).

$$Cls_{\text{Loss}} = -\sum_{i=1}^{M} [\omega_j y_j \log(p_j) + (1 - y_j) \log(1 - p_j)] \tag{18}$$

Distribution Focal Loss (Dfl_{Loss}) is utilized to improve the prediction of bounding boxes, and it is focused on the bounding box's precise localization (Behzadi *et al.*, 2025). This method enables the model to concentrate not only on enhancing overall localization but also on obtaining finer information in areas of high confidence. Consequently, it enhances bounding box precision and regression quality (Li *et al.*, 2025). It is calculated by applying Equation (19).

$$Dfl_{\text{Loss}} = \sum_{i=1}^{M} (p_j \times X_j - X_j) \tag{19}$$

Where: **TP** is the number of samples that the model accurately classified as positive, signifying precise detections. **FP** is the number of samples that the model misclassifies as positive, resulting in erroneous warnings or misclassifications. **FN** is the quantity of real positive samples that were overlooked by the model, or missed detections. **N** is the number of object classes (Kutyrev *et al.*, 2025). **AP_i** is the class's average precision, which is determined by calculating the area under the class's precision-recall curve, encapsulating the model's equilibrium between recall and precision across recall levels (Zhu *et al.*, 2025). **P(r)** stands for the precision rate that, in the precision–recall rate curve (P–R Curve), corresponds to the recall rate (Zhang *et al.*, 2025). **Euclidean distance** between the center points of the boxes (ground-truth box (b) and predicted box (b^*)), measures the deviation of the central point (the height and width of the ground truth and predicted bounding boxes series). **C** is the minimal enclosing bounding box's diagonal distance. **V** stands for aspect ratio consistency term (penalizes the discrepancy between the ratio of predicted boxes to ground truth). α represents the correction factor.

M is the number of classes. $\mathbf{y_j}$ is the ground truth sample j. $\mathbf{p_j}$ represents the expected probability associated with sample j. ω_i is the weight of the predicted and ground truth bounding boxes' coordinates (Behzadi *et al.*, 2025).

5. RESULTS

The main results, which aimed at enhancing the detection of all dental diseases using the YOLO11 with the integrated dataset, are presented in this section. The trained model performance was checked using the validation sets of precision (P), recall (R), mean average precision (mAP) when intersection over union (IOU) is greater than or equal 0.5 (mAP@50), mAP when intersection over union (IOU) is greater than or equal 0.5 and less than 0.95 (mAP@50-95) and F1-score.

It is important to notice that the first row of the results in the tables below shows the overall achievement in each test. The overall results are calculated using Equation (20), for M to be the number of classes:

$$Overall = \frac{\text{result of row\#2} + \text{result of row\#3} + \cdots + \text{result of row\#M}}{M} \tag{20}$$

Training#1: The YOLO11m model was trained on the DRD. The results are shown in Table 1 and Figures (11 and 12).

Table 1: Validation results of the trained YOLO11m on DRD.

Class	Precision	Recall	mAP@50	mAP@50-95	F1-score
Overall	0.788	0.793	0.837	0.547	0.79
Filing	0.886	0.863	0.926	0.627	-
Impacted	0.838	0.792	0.908	0.557	-
Implant	0.950	0.958	0.978	0.679	-
Cavity	0.476	0.558	0.535	0.327	-

Table 1 shows the results of training YOLO11 on the DRD. The model established robust overall performance with mAP@50 of 0.837, recall of 0.793, precision of 0.788 and F1-score of 0.79. The implant class produced the highest mAP@50 of 0.978, recall of 0.958 and precision of 0.95. The filling class achieved high mAP@50 of 0.926, with balanced precision of 0.886 and recall of 0.863. The impacted class obtained lower mAP@50 of 0.908, recall of 0.792 and precision of 0.838 as compared to the filling and implant class. The cavity class produced the lowest mAP@50 of 0.535, recall of 0.558 and precision of 0.476. This shows that the implant class achieved the maximum detection and correct classification. However, the cavity class showed the lowest correct categorization, with the highest misdetection (no detection). Overall, the model performed well in every class, with the exception of the cavity class, which is still significantly challenging.

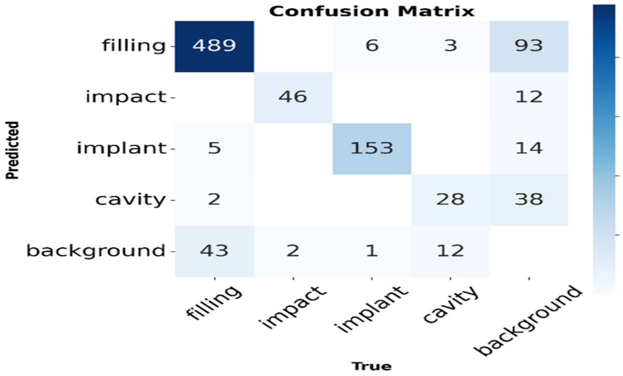


Figure 11: The confusion matrix of the DRD.

Figure 11 shows the confusion matrix of the YOLO11 with the DRD containing four classes: fillings, impacted, implant and cavity. For the filling class, the validation set contains 539 instances. The model correctly classified 489 instances, misclassified 7 instances (5 implants and 2 cavity), failed to detect 43 instances, and produced 93 instances detections without corresponding ground truth. And for the impacted class, the validation set contains 48 instances. Among them, 46 instances were correctly detected, and 2 instances were not detected. Additionally, the model produced 12 instance detections for which no ground truth was available. Also, for the implant class, the validation set contains 160 instances. Between these, 153 instances were correctly classified, 6 instances were misclassified as fillings, and 1 instance was not detected. The model produced 14 instance detections for which no ground truth was available. Moreover, the validation set of the cavity class contains 43 instances. The model correctly detected and classified 28 instances, misclassified 3 as fillings, failed to detect 12 instances, and produced 38 instance detections without corresponding ground truth. The highest correct classification was in impacted and implant, and the lowest was in cavity. The lowest missing instance was in implant, and the highest was in cavity class. This indicates that the cavity has the highest misclassification and misdetection, and the detection of the cavity class using panoramic X-ray to be difficult.

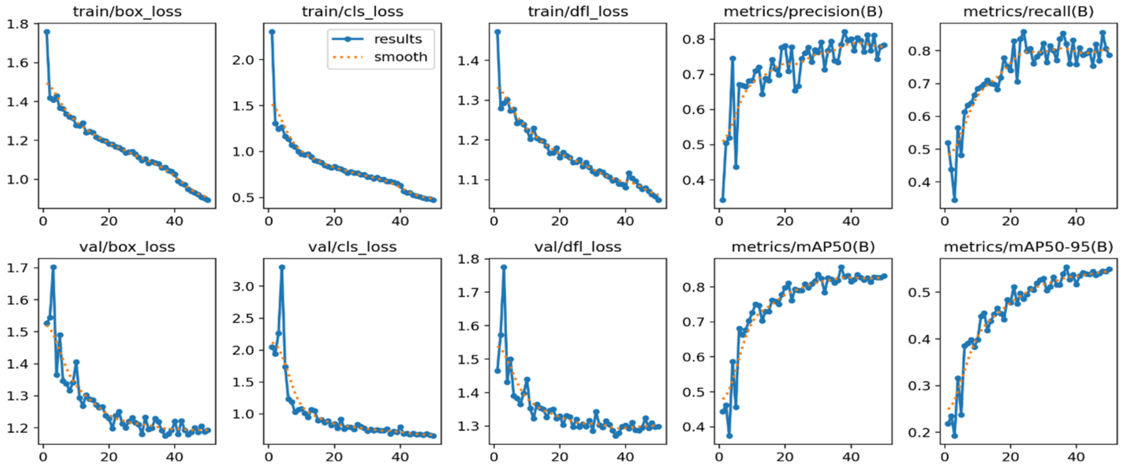


Figure 12: The curve of the loss type for validation and training, precision, recall, mAP@50, and mAP@50-95 of the DRD.

Figure 12 shows performance findings of the YOLO11 model on the DRD. The **precision** and recall rates fluctuated and increased over the epochs, demonstrating better generalization and learning. mAP@50 and mAP@50-95 gradually increased, while train and validation (box loss, classification loss, distribution focal loss) decreased, indicating an enhanced model and no overfitting.

Training#2: Training#1 is repeated on the DRD without the cavity class. The outcomes are shown in Table 2 and Figures (13 and 14).

Table 2: Validation results of the trained YOLO11m on DRD with the cavity class removed.

Class	Precision	Recall	mAP@50	mAP@50-95	F1-score
Overall	0.847	0.926	0.936	0.618	0.88
Filing	0.822	0.907	0.924	0.615	
Impact	0.813	0.906	0.919	0.569	
Implant	0.906	0.965	0.965	0.670	

Table 2 presents the results of the first training when repeated on the DRD with the excluded cavity class. The model showed robust and increased overall performance in mAP@50 of 0.936, recall of 0.926, precision of 0.847 and F1-score of 0.88. The implant class yielded the highest precision of 0.906, recall of 0.965, and mAP@50 of 0.965. The filling class achieved high mAP@50 of 0.924, precision of 0.822 and recall of 0.907. The impacted class obtained lower mAP@50 of 0.919, recall of 0.906 and precision of 0.813 compared to the filling and implant class. Overall, the model performed excellently in correct classification and detection, and low- rate misdetection.

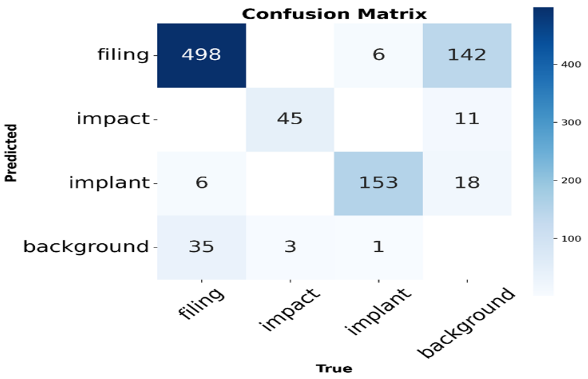


Figure 13: The confusion matrix of the DRD without cavity.

Figure 13 presents the confusion matrix results of YOLO11 with the DRD when the cavity class removed. After the number of classes decreased, the performance of the filling class slightly increased, the correct prediction increased to 498 instances, and misclassification and misdetection reduced to 6 and 35 instances. Moreover, detection without corresponding ground truth increased to 142 instances. Additionally, the correct classification of the impacted class reduced to 45 instances, with misdetection increased to 3 instances, and detection without corresponding ground truth decreased to 11 instances. However, the implant class remained the same without changing, only the detection without corresponding ground truth decreased to 18 instances. The correct classification of all classes is excellent, and misclassification and misdetection is very low; this indicates a higher performance of the model.

Figure 14 illustrates the performance of YOLO11 model on the DRD with the excluded cavity class. Compared to the first training, precision and recall rose above 0.8, indicating lower misclassification and higher detection cases. Additionally, mAP@50 and mAP@50-95 increased, while box loss, classification loss, and distribution focal loss decreased. These results indicate better generalization, enhancement, and detection.

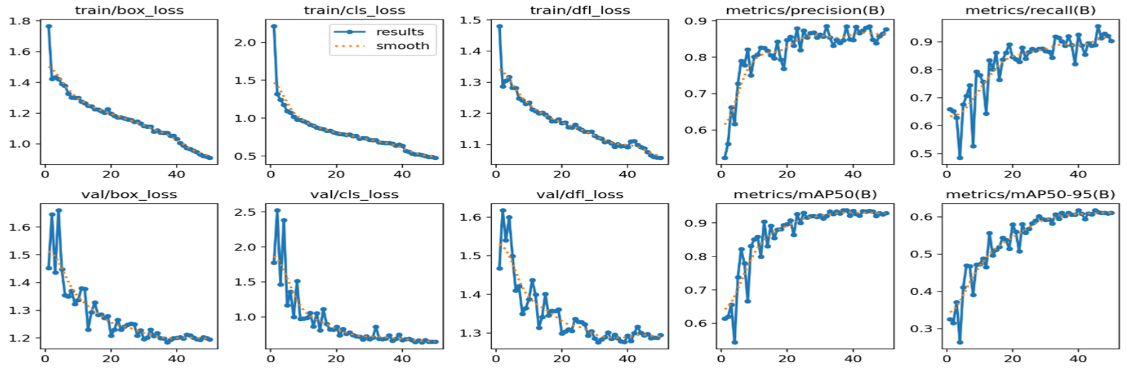


Figure 14: The curve of the loss type for validation and training, precision, recall, mAP@50, and mAP@50-95 of the DRD without cavity.

Table 3: Validation results of the trained YOLO1m on the RID.

Class	Precision	Recall	mAP@50	mAP@50-95	F1-score
Overall	0.764	0.838	0.858	0.754	0.8
Decay cavity	0.873	0.868	0.910	0.788	
Early decay	0.684	0.887	0.863	0.766	
Healthy teeth	0.736	0.759	0.802	0.707	

Table 3 demonstrates the results of training YOLO11 on the RID. The model achieved powerful overall performance, mAP@50 of 0.858, recall of 0.838, precision of 0.764 and F1-score of 0.8. The decay cavity exhibited the best metrics through all classes, with mAP@50 of 0.91, balanced recall of 0.868 and precision of 0.873. The early cavity class achieved strong mAP@50 of 0.863, and recall of 0.887 with low precision of 0.684. The healthy teeth produced high mAP@50 of 0.802, with balanced recall of 0.759 and precision of 0.736. The image type in this test was colorful Red-Green-Blue (RGB), that depended on the color of the teeth for classification. This indicates higher resolution images due to lower noise in capturing, with the setting and type of the device, lighting and angle of capturing (as the shadow affects the classification). For this reason, the classification of the early decay is hard, because the same healthy teeth misclassified with other unhealthy teeth significantly. In general, the model performed well over all classes in classification and detection, with correct prediction and low rate of misdetection.

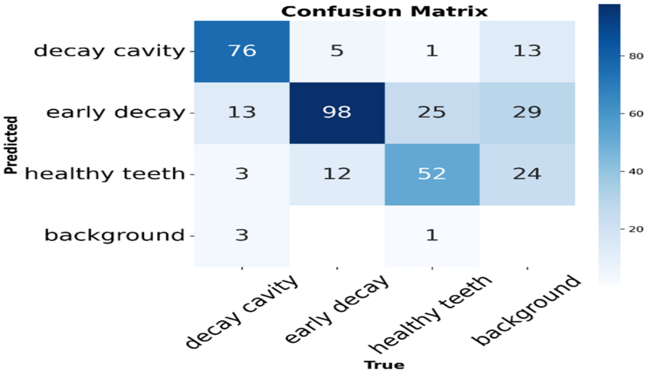


Figure 15: The confusion matrix of the RID.

Figure 15 summarizes the confusion matrix of YOLO11 with the RID, which includes three classes: decay cavity, early cavity, and healthy teeth. The early cavity contains 115 instances in the validation set, among them YOLO correctly classified 98 instances, with 17 instances confused with healthy teeth and decay cavity, and produced 29 detection instances when they were unavailable in the ground truth. The decay cavity in the validation set contains 95 cases, 76 were correctly detected, 13 confused as early cavity, 3 as healthy teeth, and also 13 instances were detected without corresponding ground truth, and 3 instances were not predicted by the model. Additionally, the healthy teeth were 79 instances out of them 52 instances correctly classified, 1 instance misclassified with the decay cavity and 25 as early cavities, and 24 cases were detected without corresponding ground truth. The highest correctly classified class was early cavity and the lowest was healthy teeth. The lowest miss detected class was healthy teeth and the highest decay cavity. The performance of the detection and correctly classification is countered as good.

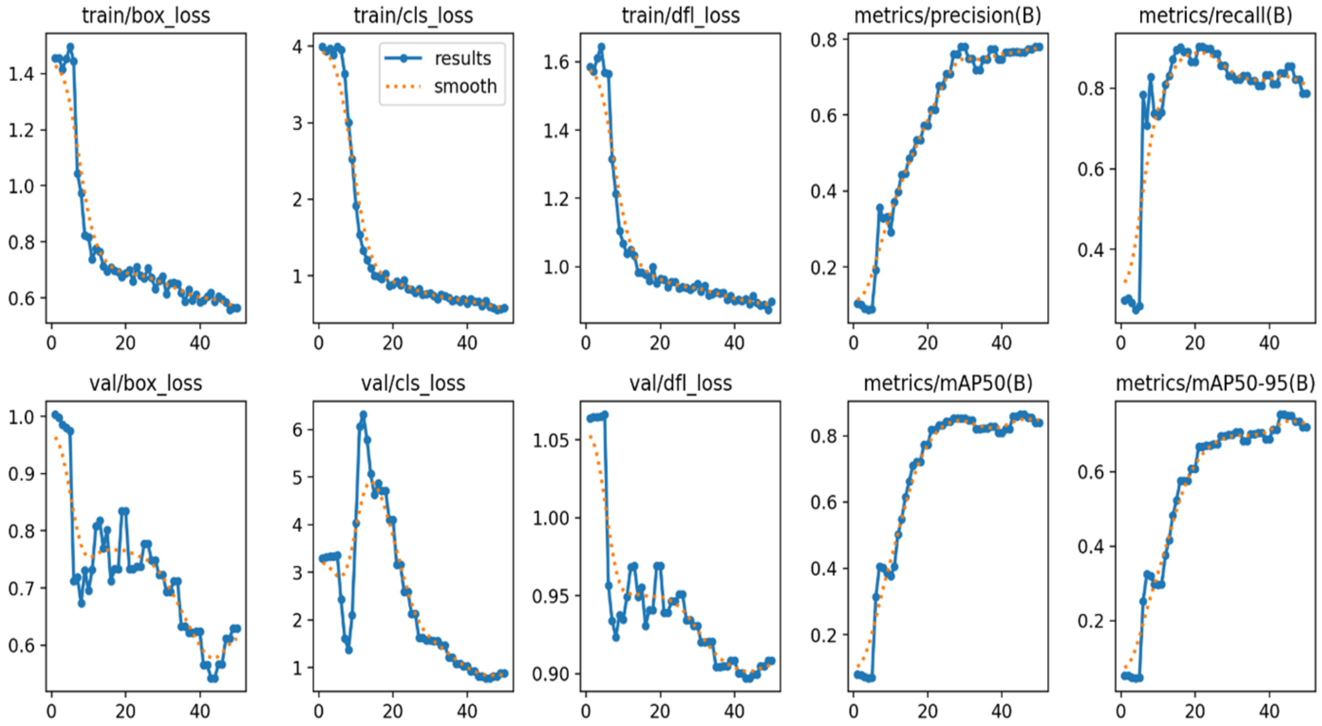


Figure 16: The curve of the loss type for validation and training, precision, recall, mAP@50, and mAP@50-95 of the RID.

Figure 16 demonstrates the performance of evaluation metrics on the RID for YOLO11 model. Initially, Precision and Recall increased, then stabilized, indicating early learning and accurate prediction and recall. However, mAP@50 and mAP@50-95 were gradually increased, while some of training and validation (box loss, classification loss, distribution focal loss) fluctuated but all decreased over the epochs, which signifies an enhanced model with high efficiency and no overfitting.

Training#4: More than 26 attempts were applied on the RID with size of 80 × 80 pixels, followed by a combination of the 64 images were put on one frame of size 640 × 640 pixels. The combination and the RID were both trained with YOLO11m in each attempt as shown in Table 4. The highest performing augmentation was selected, the outcome demonstrated in detail in Table 5 and Figures (17 and 18).

Table 4: Examples of the training by adding collection of augmentation techniques on the RID.

#	Technique/s and Range	mAP@50
1	Original	0.858
2	Rotate (CW)	0.850
3	Rotate (CCW)	0.861
4	Shear ±10° H/V	0.856
5	Crop 0%–20% Zoom	0.846
6	Brightness ±15%	0.845
7	Flip (H, V), Rotate (CW, CCW, UD), Zoom 0–20%, Shear ±10° H/V, Brightness ±15%	0.862
8	Rotate (CW, CCW, UD), Shear ±10° H/V, Saturation ±25%	0.882
9	Flip (H, V), Rotate (CW, CCW, UD), Shear ±10° H/V, Saturation ±25%	0.842
10	Rotate (CW, CCW, UD), Zoom 0%–20%, Shear ±10° H/V, Brightness ±15%	0.850
11	Flip (H, V), Rotate (CW, CCW, UD), Zoom 0%–20%, Brightness ±15%	0.844
12	Flip (H), Rotate (CW, CCW, UD), Zoom 0%–20%	0.899
13	Flip (V), Rotate (CW, CCW, UD), Shear ±10° H/V, Saturation ±25%	0.867
14	Rotate (CW, CCW), Shear ±10° H/V, Saturation ±25%	0.836
15	Rotate (CW), Shear ±10° H/V, Saturation ±25%	0.865
16	Rotate (CCW, UD), Shear ±10° H/V, Saturation ±25%	0.860
17	Rotate (CCW), Shear ±10° H/V, Saturation ±25%	0.860
18	Rotate (UD), Shear ±10° H/V, Saturation ±25%	0.866
19	Flip (H), Rotate (CW, CCW, UD), Zoom 0%–20%	0.854
20	Flip (H), Rotate (CW, CCW, UD)	0.876
21	Flip (H), Zoom 0%–20%	0.854
22	Flip (H), Rotate (UD), Zoom 0%–20%	0.880
23	Flip (H), Rotate (CW, CCW), Zoom 0%–20%	0.847
24	Flip (H), Rotate (CCW, UD), Zoom 0%–20%	0.868
25	Flip (H), Rotate (CW), Zoom 0%–20%	0.859
26	Flip (H), Rotate (CW), Zoom 0%–20%	0.864

Where: H represents Horizontal Flip. V represents Vertical Flip. CW represents Clock Wise. CCW represents Counters Clock Wise. UD represents Upside Down (i.e. 180° Rotation).

Table 4 presents several collections of augmentations applied to the RID for enhancing detection accuracy. Many augmentation techniques, such as rotate, flip, zoom, shear, saturated, and brightness, were utilized to find the best enhancement in the RID. Rotation consists of three options 90 clock wise, 90 counter clock wise, and 180 upsides down. Flip provides two options; horizontal and vertical. Saturated and brightness make the image 10% lighter or darker. Zoom crops the image from (0 to 20%) then resizes it to 80*80 pixels. After more than 26 changes of the type and the option of the augmentation techniques as described in table, the 12th training produced the highest mAP@50 of 0.899 with collection augmentation 90° clockwise, 90° counter clockwise, 180° upsides down, horizontal flip, zoom up to 20% randomly. The augmentation enhanced detection and classification when the suitable type applied, because adding diversity and variations to the dataset were without effort of collecting new images. These techniques also prevented the model from overfitting.

Table 5: Validation results of the trained YOLO11m on the augmented RID.

Class	Precision	Recall	mAP@50	mAP@50-95	F1-score
Overall	0.787	0.888	0.899	0.784	0.83
Decay cavity	0.931	0.849	0.924	0.790	
Early decay	0.752	0.852	0.897	0.797	
Healthy teeth	0.672	0.962	0.877	0.764	

Table 5 displays the results of training YOLO11 on the augmented RID. The model improved overall performance after augmentation compared to the raw RID achieving mAP@50 of 0.899, precision of 0.787, recall of 0.888, and F1-score of 0.83. The most enhanced class across all classes was the healthy teeth with mAP@50 of 0.877 which improved about 7.5than 20for 0.752 improvement (about 6.8The decay cavity provided the best metrics, with mAP@50 of 0.924, precision of 0.931, high recall of 0.849. In addition to enhancement in mAP@50 and precision, the model's performance was enhanced across all classes in classification and detection, demonstrating correct prediction and low rate misdetections.

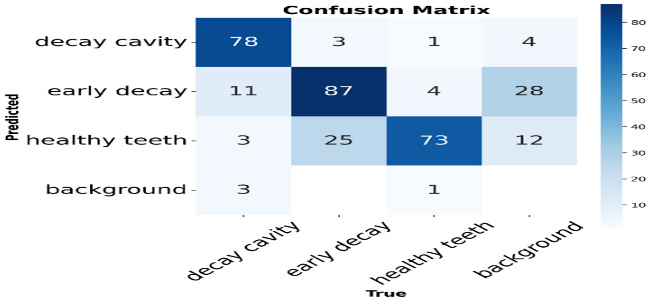


Figure 17: The confusion matrix of the augmented RID.

Figure 17 presents the enhanced confusion matrix of the augmented RID with the utilized model. After enhancing the decay cavity, the correct classification increased to 78, with decrease in misclassifications to 14 and detection without corresponding ground truth to 4. Additionally, the correct prediction in healthy teeth increased to 73, misclassification significantly decreased to 5, other things remain the same. For early cavity, after augmentation correctly 87 images were predicted and detection without corresponding ground truth reduced to 28, with increment in confusion.

Figure 18 presents performance results of YOLO11 on the augmented RID. The Precision and Recall curves increased compared to training#3, indicating fewer misclassifications and enhanced detection. Meanwhile, mAP@50 and mAP@50-95 rapidly increased at the beginning, then steadily increased and got higher than the raw RID, indicating improved object recognition. Training and validation (box loss, classification loss, distribution focal loss) decreased over the time, demonstrating no overfitting.

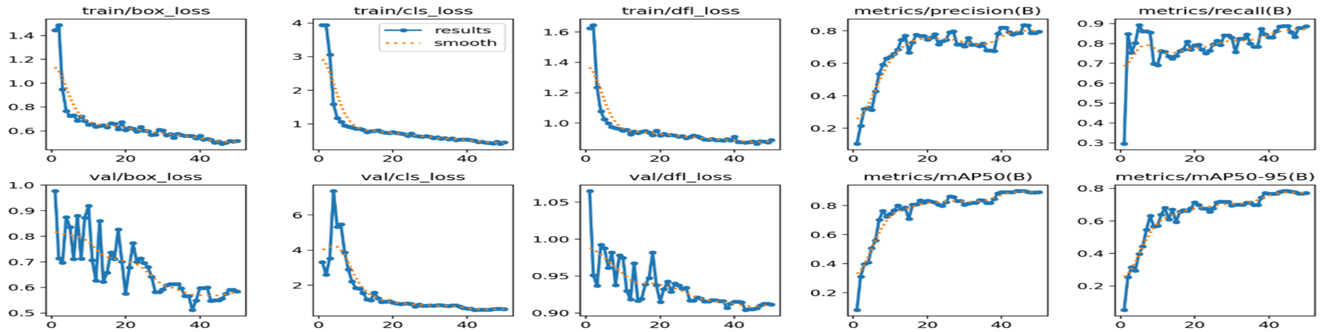


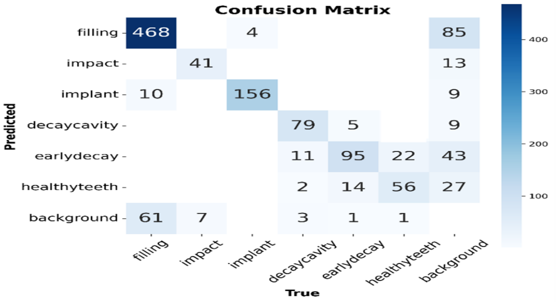
Figure 18: The loss type curves for validation and training, precision, recall, mAP@50, and mAP@50-95 of the augmented RID.

Training#5: The integrated dataset trained on YOLO11m. The results are presented in Table 6 and Figures (19 and 20).

Table 6: Validation results of trained YOLO11m on the integrated dataset.

Class	Precision	Recall	mAP@50	mAP@50-95	F1-score
Overall	0.813	0.883	0.904	0.691	0.84
Filing	0.877	0.865	0.927	0.620	
Impact teeth	0.783	0.854	0.903	0.563	
Implant	0.921	0.975	0.978	0.669	
Decay cavity	0.922	0.849	0.912	0.770	
Early decay	0.651	0.893	0.903	0.807	
Healthy teeth	0.724	0.864	0.802	0.718	

Table 6 illustrates the results of training YOLO11 on the integrated dataset. The model demonstrated tough overall performance with mAP@50 of 0.904, precision of 0.813, recall of 0.883 and F1-score of 0.84. Among all classes the implant achieved the highest mAP@50 of 0.978, recall of 0.975 and high precision of 0.921. Additionally, the filling class produced robust mAP@50 of 0.927, with balanced recall of 0.865 and precision of 0.877. Also, the early decay acquired excellent mAP@50 of 0.903, recall of 0.893, and presented low precision of 0.651. Across all classes the decay cavity class achieved the highest precision of 0.922, with high recall of 0.849 and mAP@50 of 0.912. The healthy teeth found moderated with mAP@50 of 0.802, precision of 0.724, and recall of 0.864. Overall, the model's performance was perfect for all classes, with significant enhancement in classification and detection compared to the first training. Additionally, the early cavity was added for more detail as it is a very important class to prevent from cavity or bone losses.



	filling	impact	implant	decaycavity	earlydecay	healthyteeth	background
filling	468		4				85
impact		41					13
implant	10		156				9
decaycavity				79	5		9
earlydecay				11	95	22	43
healthyteeth				2	14	56	27
background	61	7		3	1	1	

Figure 19: The confusion matrix of the integrated dataset.

Figure 19 demonstrates the confusion matrix for six classes (filling, impacted, implant, cavity decay, early decay, and healthy teeth), evaluating the performance of the integrated dataset with the YOLO11. The validation set consists of 1036 samples. 539 instances annotated as fillings, 468 instances correctly categorized, with 10 instances confused as implants, and 85 instance were detected without corresponding ground truth. Out of 160 instances in the implant category, 156 instances were correctly identified, 4 instances were misclassified with the filling class, and 9 instances were detected without corresponding ground truth.

115 instances annotated as the early decay, 95 correctly predicted, with misclassified 19 instances as healthy teeth and decay cavity, and 43 instances detected without corresponding ground truth. From 95 samples of decay cavity, 79 instances were correctly detected, 13 instances were misclassified (11 instances as early decay and 2 instances as healthy teeth), with 3 undetected samples, and 9 instances detected without corresponding ground truth. For healthy teeth, the model correctly classified 56 instances, 22 instances confused with early decay, 1 sample was not detected, and 27 instances were detected without corresponding ground truth. Among 48 impact instances, 41 instances were correctly classified, 7 instances were not detected, and 13 instances were detected without corresponding ground truth. The correct classification and detection of all classes were perfect, some of them decreased, because the number of classes increased.

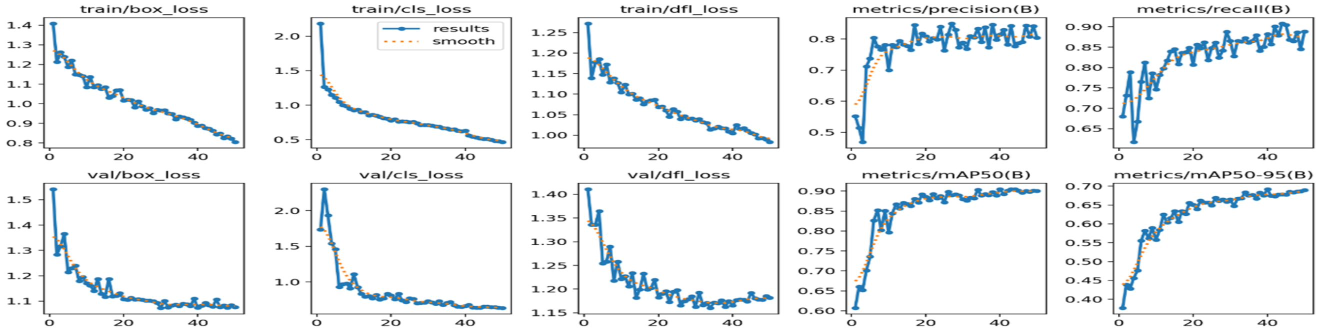


Figure 20: The loss type curves for validation and training, precision, recall, mAP@50, and mAP@50-95 of the integrated dataset.

Figure 20 determines the performance of the result of YOLO11 on the integrated dataset. At first, the precision curves rapidly rose, then stabilized, also the recall gradually increased, which signifies fewer missing instances and high classification and detection. Meanwhile, mAP@50 and mAP@50-95 increased over time, training and validation (box loss, classification loss, distribution focal loss) decreased over time, signifying no overfitting. Overall, the results indicated high generalization, detection, and classification with fewer confusing instances as compared to the first training, with more detailed disease detection.

6. COMPARISON

The mAP@50 evaluation for the system has been chosen in this work as it is the most comprehensive and widely used evaluation metric due to its capability to function as an accuracy measure in object detection (Zaidi *et al.*, 2022). To date, this assessment metric is the most powerful approach to calculate the accuracy which depends on precision, recall, and the number of classes (Zhang *et al.*, 2025). A greater mAP@50 score reflects stronger model performance in identifying and categorizing disease in images (Du, 2023).

A comprehensive review of YOLO model variants for dental disease detection has been conducted in recent literature. The review is shown in Figure 15 as a chart. Alkentar and Assalem (2024) evaluated YOLOv7, YOLOv8, and YOLOv9 on the DRD containing 1269 images, reporting mAP@50 scores of 40.3%, 76.6%, and 75.9%, respectively. Similarly, Thulaseedharan and PS (2022, 2023) trained YOLOv5, YOLOv6 models on private datasets containing 664 images, and they achieved mAP@50 values of 72.4% and 70.76%, respectively. Also, Maged *et al.* (2024) and George *et al.* (2023) performed YOLOv8 on the same dataset (DRD) using different methods for improving mAP@50, the results were (80.6% and 82.1%) respectively. George *et al.* (2023) segmented the images three times before training; whereas, Maged *et al.* (2024) directly trained on the dataset. In comparison of the results in this work, it is noticed that the adaptation of YOLO11 for DRD execution enhanced the detection by more than 1.6% which shows improvements achieved by this model compared to other versions and more than 14.5% higher mAP@50 is accomplished through the integrated dataset.

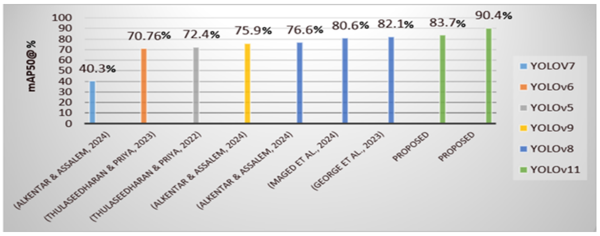


Figure 21: . Comparison between the results of the proposed work and previous studies.

7. DISCUSSION

In recent years, automatic detection in medical images has rapidly grown and attracted the interest of researchers. However, the quality and resolution of medical images remained limited. This paper aimed to present a system capable of accurately detecting the most significant dental diseases, which are cavities, fillings, impacted teeth, and implant teeth.

According to the results achieved in this work, training YOLO 11 on the DRD improved the overall detection accuracy. Nevertheless, the accuracy for cavity detection remained relatively low. Training YOLO11 on the RID in cavity class gave much higher detection than panoramic X-rays; however, detection of impacted and implanted teeth in real images was significantly hard. It is observed that each class can be detected accurately in a specific type of image. For instant, it is recommended to use the RID for the cavity class over the panoramic X-rays. On the contrary, detection of impacted and implant teeth diseases in real images is relatively complicated. Furthermore, in such cases, it is often necessary to use two separate application systems; one specialized for cavity detection and another for implant and impacted diseases to achieve higher diagnostic accuracy. To overcome diagnosis issues and improve the detection of all dental diseases in a model, a new dataset has been generated by integrating a panoramic X-ray image dataset with a real image dataset. The integration of two datasets can be considered a preprocessing step. The deep learning (DL) model is trained on the new dataset for accurate detection of all dental diseases. The integration of two datasets is to leverage their complementary strengths, resulting in a unified and more robust model that improves detection performance across all classes, with a notable enhancement in cavity detection.

The research limitations and challenges of this work were the low resolution of DRD images, which reduced overall detection accuracy. Also, the imbalance in the number of classes in the dataset, as sometimes the same image contains multiple diseases, and when a class increased, another class is also increased. This reduced the detection accuracy of the system. Moreover, the availability of a limited number of datasets having a limited number of image samples.

8. CONCLUSIONS

Dental diseases are the major global health concerns. Dental caries is one of the most common dental diseases which can lead to infection and tooth loss, making early detection critical for prevention and treatment. The proposed integrated dataset demonstrated a significant enhancement in detection. In this study it is observed that:

- In the first training, the cavity class in the DRD has a lower detection rate, resulting in a decline in the accuracy of the model overall.
- In the second training, removing the cavity class significantly increased the mAP@50 of the overall achievements
- In the third training, the cavity class in the real image gave much higher detection and mAP@50 than the cavity class in the panoramic x-ray image.
- In the fourth training, adding augmentation to the RID caused higher detection.
- In the fifth training, integrating two datasets (DRD without cavity class + RID) enhanced the detection of all classes and the system's overall performance.

Future improvements will aim to develop a comprehensive dataset capable of detecting a wide range of dental conditions such as missing teeth and bridge restorations, oral diseases such as cysts and oral cancer, and decorative or cosmetic issues such as cracks, discoloration, and misalignment. The objective is to enable a system not to diagnose only clinical problems but also to assess aesthetic dental concerns. Additionally, future studies will focus on optimizing the algorithm to reduce the computational complexity by excluding redundant network blocks or employing model compression techniques without compromising detection accuracy. Furthermore, a user-friendly mobile application will be developed to allow individuals to perform basic dental assessments at home, improving accessibility to preliminary oral health screening.

Finally, it is important to mention that the proposed model can be readily implemented in dental clinics and also holds strong potential for integration into a mobile application. In such an application, users could capture intraoral images using a smartphone and upload them to detect specific dental conditions, such as cavity types. Additionally, users or practitioners could upload panoramic X-ray images to verify diagnostic results through automated analysis.

Acknowledgement

This work would not have been possible without the collective efforts and encouragement of the Faculty of Engineering at University of Duhok to whom we remain deeply indebted.

Ethical statement

This work would not have been possible without the collective efforts and encouragement of the Faculty of Engineering at University of Duhok to whom we remain deeply indebted.

Funding

We wish to clarify that this work did not receive any funding from external organizations, agencies, or institutions.

Author Contributions

M.A.A. and M.A.S. contributed equally to the conception and design of the study. M.A.A. performed data acquisition, analysis, or interpretation, and drafting of the manuscript. Both authors have read and agreed to the published version of the manuscript.

9. REFERENCES

Ahmed, S., Saifullah, Ahmad, M., Swami, B. L., & Ikram, S. (2016). Green synthesis of silver nanoparticles using Azadirachta indica aqueous leaf extract. *Journal of Radiation Research and Applied Sciences*, 9(1), 1–7. DOI:https://doi.org/10.1016/j.jrras.2015.06.006

Alibe, I. M., Matori, A. K., Saion, E., Ali, A. M., & Zaid, M. H. M. (2017). The influence of calcination temperature on structural and optical properties of ZnO nanoparticles via simple polymer synthesis route. *Science of Sintering*, 49(3), 263–275. DOI:https://doi.org/10.2298/SOS1703263A

Altammar, K. A. (2023). A review on nanoparticles: characteristics, synthesis, applications, and challenges. *Frontiers in Microbiology*, 14, 1155622. DOI:https://doi.org/10.3389/fmicb.2023.1155622

Alzoubi, F., Ahmad, A. A., Aljarrah, I. A., Migdadi, A., & Al-Bataineh, Q. M. (2023). Localize surface plasmon resonance of silver nanoparticles using Mie theory. *Journal of Materials Science: Materials in Electronics*, 34(32), 2128. DOI:https://doi.org/10.1007/s10854-023-11304-x

Barzinjy, A. A., & Azeez, H. H. (2020). Green synthesis and characterization of zinc oxide nanoparticles using Eucalyptus globulus Labill. leaf extract and zinc nitrate hexahydrate salt. *SN Applied Sciences*, 2(5), 991. DOI:https://doi.org/10.1007/s42452-020-2813-1

Barzinjy, A. A., Hamad, S. M., Abdulrahman, A. F., Biro, S. J., & Ghafor, A. A. (2020). Biosynthesis, characterization and mechanism of formation of ZnO nanoparticles using Petroselinum crispum leaf extract. *Current Organic Synthesis*, 17(7), 558–566. DOI:https://doi.org/10.2174/1570179417666200628140547

Bekele, B., Jule, L., & Saka, A. (2021). The effects of annealing temperature on size, shape, structure and optical properties of synthesized zinc oxide nanoparticles by sol-gel methods. *Digest Journal of Nanomaterials and Biostructures*, 16(2), 471–478. DOI:http://dx.doi.org/10.15251/DJNB.2021.162.471

Bian, S.-W., Mudunkotuwa, I., Rupasinghe, T., & Grassian, V. (2011). Aggregation and Dissolution of 4 nm ZnO Nanoparticles in Aqueous Environments: Influence of pH, Ionic Strength, Size, and Adsorption of Humic Acid. *Langmuir*, 27, 6059–6068. DOI:http://dx.doi.org/10.1021/la200570n

Bolhassani, A. (2023). Lipid-based delivery systems in development of genetic and subunit vaccines. *Molecular Biotechnology*, 65(5), 669–698. DOI:https://doi.org/10.1007/s12033-022-00624-8

Burg, D., & Ausubel, J. H. (2021). Moore's Law revisited through Intel chip density. *PLoS ONE*, 16(8), e0256245. DOI:https://doi.org/10.1371/journal.pone.0256245

Chang, J. S., Saint, C. P., Chow, C. W., Bahnemann, D. W., & Chong, M. N. (2024). Recent innovations in engineering Zinc Oxide (ZnO) nanostructures for water and wastewater treatment: Pushing the boundaries of multifunctional photocatalytic and advanced biotechnological applications. *International Materials Reviews*, 69(7-8), 337–379. DOI:https://doi.org/10.1177/09506608241280421

ElFaham, M. M., Mostafa, A. M., & Mwafy, E. A. (2021). The effect of reaction temperature on structural, optical and electrical properties of tunable ZnO nanoparticles synthesized by hydrothermal method. *Journal of Physics and Chemistry of Solids*, 154, 110089. DOI:https://doi.org/10.1016/j.jpcs.2021.110089

Fadaie, N., Sheikhi, M., Mohammadi, T., Tofighy, M. A., Rajabzadeh, S., & Sahebi, S. (2021). Novel Plasma Functionalized Graphene Nanoplatelets (GNPs) incorporated in forward osmosis substrate with improved performance and tensile strength. *Journal of Environmental Chemical Engineering*, 9(4), 105708. DOI:https://doi.org/10.1016/j.jece.2021.105708

Gopinath, S. C. (2020). *Nanoparticles in Analytical and Medical Devices*. Elsevier. DOI:https://doi.org/10.1016/B978-0-12-821163-2.00001-7

Holzwarth, U., & Gibson, N. (2011). The Scherrer equation versus the 'Debye-Scherrer equation'. *Nature Nanotechnology*, 6(9), 534–534. DOI:https://doi.org/10.1038/nnano.2011.145

Ismail, A. M., Menazea, A., Kabary, H. A., El-Sherbiny, A., & Samy, A. (2019). The influence of calcination temperature on structural and antimicrobial characteristics of zinc oxide nanoparticles synthesized by Sol–Gel method. *Journal of Molecular Structure*, 1196, 332–337. DOI:https://doi.org/10.1016/j.molstruc.2019.06.084

Ismail, S. M., & Ahmed, S. M. (2023). The effect of calcination temperatures on the properties of ZnO nanoparticles synthesized by using leaves extracts of pinus brutia tree. *Science Journal of University of Zakho*, 11(2), 286–297. DOI:https://doi.org/10.25271/sjuoz.2023.11.2.1087

Kaur, A., Kaur, H., Singh, R., & Singh, V. K. (2025). Implications on Structural, Morphological, and Optical Properties and Photocatalytic Degradation Behavior of Zinc Oxide Nanoparticles: Effect of Calcination Temperature. *JOM*, 1–11. DOI:https://doi.org/10.1007/s11837-025-07344-9

Kayaci, F., Vempati, S., Donmez, I., Biyikli, N., & Uyar, T. (2014). Role of zinc interstitials and oxygen vacancies of ZnO in photocatalysis: A bottom-up approach to control defect density. *Nanoscale*, 6(17), 10224–10234. DOI:http://dx.doi.org/10.1039/c4nr01887g

Khan, S., & Hossain, M. K. (2022). Classification and properties of nanoparticles. In *Nanoparticle-Based Polymer Composites* (pp. 15–54). Elsevier. DOI:https://doi.org/10.1016/B978-0-12-824272-8.00009-9

Kotousov, A., Khanna, A., & Bun, S. (2014). An analysis of elastoplastic fracture criteria. In L. Ye (Ed.), *Recent Advances in Structural Integrity Analysis - Proceedings of the International Congress (APCF/SIF-2014)* (pp. 67–71). Woodhead Publishing. DOI:https://doi.org/10.1533/9780081002254.67

Kumar, S., Hussain, A., Siddiqui, A. M., Khan, Z. H., Abdullah, M. M., & Ashraf, M. T. (2024). Synthesis and study of the impact of calcination duration on the properties of Al4(ZnO)96 nanoparticles. *Nano-Structures & Nano-Objects*, 39, 101250. DOI:https://doi.org/10.1016/j.nanoso.2024.101250

Loza, K., Epple, M., & Maskos, M. (2019). Stability of nanoparticle dispersions and particle agglomeration. *Biological Responses to Nanoscale Particles: Molecular and Cellular Aspects, Method-*

ological Approaches, 85–100. DOI:https://doi.org/10.1007/978-3-030-12461-8_4

Mahajan, M., Kumar, S., Gaur, J., Kaushal, S., Dalal, J., Singh, G., ... & Ahlawat, D. S. (2025). Green synthesis of ZnO nanoparticles using Justicia adhatoda for photocatalytic degradation of malachite green and reduction of 4-nitrophenol. *RSC Advances*, 15(4), 2958–2980. DOI:https://doi.org/10.1039/D4RA08632E

Mahato, S. S., Mahata, D., Panda, S., & Mahata, S. (2023). Perspective chapter: sol-gel science and technology in context of nanomaterials–recent advances. *Sol-Gel Method - Recent Advances*. DOI:https://doi.org/10.5772/intechopen.111378

Malik, S., Muhammad, K., & Waheed, Y. (2023). Nanotechnology: a revolution in modern industry. *Molecules*, 28(2), 661. DOI:https://doi.org/10.3390/molecules28020661

Mohamed, A. F., & Mawlud, S. Q. (2023). Effect of zinc oxide on structural and optical properties borotellurite glass: ternary glass. *Zanco Journal of Pure and Applied Sciences*, 35(4), 14–21. DOI:http://dx.doi.org/10.21271/ZJPAS.35.4.02

Nagar, V., Singh, T., Tiwari, Y., Aseri, V., Pandit, P. P., Chopade, R. L., ... & Awasthi, G. (2022). ZnO Nanoparticles: Exposure, toxicity mechanism and assessment. *Materials Today: Proceedings*, 69, 56–63. DOI:https://doi.org/10.1016/j.matpr.2022.09.001

Nandhini, J., Karthikeyan, E., & Rajeshkumar, S. (2024). Green synthesis of zinc oxide nanoparticles: eco-friendly advancements for biomedical marvels. *Resources Chemicals and Materials*. DOI:https://doi.org/10.1016/j.recm.2024.05.001

Onugwu, A. L., Nwagwu, C. S., Onugwu, O. S., Echezona, A. C., Agbo, C. P., Ihim, S. A., ... & Khutoryanskiy, V. V. (2023). Nanotechnology based drug delivery systems for the treatment of anterior segment eye diseases. *Journal of Controlled Release*, 354, 465–488. DOI:https://doi.org/10.1016/j.jconrel.2023.01.018

Owuna, F. J. (2020). Stability of vegetable based oils used in the formulation of ecofriendly lubricants – a review. *Egyptian Journal of Petroleum*, 29(3), 251–256. DOI:https://doi.org/10.1016/j.ejpe.2020.09.003

Pal Singh, R. P., Hudiara, I. S., & Bhushan Rana, S. (2016). Effect of calcination temperature on the structural, optical and magnetic properties of pure and Fe-doped ZnO nanoparticles. *Materials Science-Poland*, 34(2), 451–459. DOI:https://doi.org/10.1515/msp-2016-0059

Parajuli, D., Dangi, S., Sharma, B. R., Shah, N. L., & KC, D. (2023). Sol-gel synthesis, characterization of ZnO thin films on different substrates, and bandgap calculation by the Tauc plot method. *Bibechana*, 20(2), 113–125. DOI:https://doi.org/10.3126/bibechana.v20i2.54115

Roy, A., Healey, C. P., Larm, N. E., Ishtaweera, P., Roca, M., & Baker, G. A. (2024). The huge role of tiny impurities in nanoscale synthesis. *ACS Nanoscience Au*, 4(3), 176–193. DOI:https://doi.org/10.1021/acsnanoscienceau.3c00056

Sa, Y., Guo, Y., Feng, X., Wang, M., Li, P., Gao, Y., ... & Jiang, T. (2017). Are different crystallinity-index-calculating methods of hydroxyapatite efficient and consistent?. *New Journal of Chemistry*, 41(13), 5723–5731. DOI:https://doi.org/10.1039/C7NJ00803A

Samantaray, S., Mallick, P., Hung, I.-M., Moniruzzaman, M., Satpathy, S. K., & Mohanty, D. (2024). Ceramic-ceramic nanocomposite materials for energy storage applications: A review. *Journal of Energy Storage*, 99, 113330. DOI:https://doi.org/10.1016/j.est.2024.113330

Shafiee, P., Nafchi, M. R., Eskandarinezhad, S., Mahmoudi, S., & Ahmadi, E. (2021). Sol-gel zinc oxide nanoparticles: advances in synthesis and applications. *Synthesis and Sintering*, 1(4), 242–254. DOI:https://doi.org/10.53063/synsint.2021.1477

Sharma, R., Sharma, A., Chuhadiya, S., Thakur, A., Kannan, M., & Dhaka, M. (2023). Air annealing evolution to physical characteristics of Cd0.85Zn0.15Te thin films: absorber layer applications to solar cell devices. *Journal of Materials Science: Materials in Electronics*, 34(18), 1403. DOI:https://doi.org/10.1007/s10854-023-10759-2

Sharma, V., Sharma, J., Kansay, V., Sharma, V. D., Sharma, A., Kumar, S., ... & Bera, M. (2023). The effect of calcination temperatures on the structural and optical properties of zinc oxide nanoparticles and their influence on the photocatalytic degradation of leather dye. *Chemical Physics Impact*, 6, 100196. DOI:https://doi.org/10.1016/j.chphi.2023.100196

Şimşek, T., Ceylan, A., Aşkın, G. Ş., & Özcan, Ş. (2022). Band gap engineering of ZnO nanocrystallites prepared via ball-milling. *Politeknik Dergisi*, 25(1), 89–94. DOI:https://doi.org/10.2339/politeknik.647702

Sridar, R., Ramanane, U. U., & Rajasimman, M. (2018). ZnO nanoparticles – Synthesis, characterization and its application for phenol removal from synthetic and pharmaceutical industry wastewater. *Environmental Nanotechnology, Monitoring & Management*, 10, 388–393. DOI:https://doi.org/10.1016/j.enmm.2018.09.003

Sugihartono, I., Retnoningtyas, A., Rustana, C., Umiatin, Yudasari, N., Isnaeni, ... & Kurniadewi, F. (2019). The influence of calcination temperature on optical properties of ZnO nanoparticles. *AIP Conference Proceedings*.

Supin, K. K., PM, P. N., & Vasundhara, M. (2023). Enhanced photocatalytic activity in ZnO nanoparticles developed using novel Lepidagathis ananthapuramensis leaf extract. *RSC Advances*, 13(3), 1497–1515. DOI:https://doi.org/10.1039/D2RA06967A

Tripathy, N., & Kim, D.-H. (2018). Metal oxide modified ZnO nanomaterials for biosensor applications. *Nano Convergence*, 5(1), 27. DOI:https://doi.org/10.1186/s40580-018-0159-9

Usman, A., Aris, A., Labaran, B., Darwish, M., & Jagaba, A. (2022). Effect of calcination temperature on the morphology, crystallinity, and photocatalytic activity of ZnO/TiO2 in selenite photoreduction from aqueous phase. *Journal of New Materials for Electrochemical Systems*, 25(4), 251–258. DOI:https://doi.org/10.14447/jnmes.v25i4.a05

Zubair, M. A., Mouri, T. K., & Chowdhury, M. T. (2023). Influence of Cu induced crystallographic disorder on the optical and lattice vibrational properties of ZnO nanoparticles. *Physical Chemistry Chemical Physics*, 25(40), 27628–27653. DOI:https://doi.org/10.1039/D3CP02015K