# OralVis Tooth Numbering - Model Evaluation Report

## 1. Confusion Matrix

## 

## 2. Key Metrics

|  |  |
| --- | --- |
| Metric | Value |
| Precision | 88.5% |
| Recall | 89.2% |
| mAP@50 | 91.6% |
| mAP@50-95 | 62.1% |

## 3. Summary of Approach

**Dataset Preparation**

The dataset was provided as a ZIP file containing images and labels folders. After extracting, the data was split into **train, validation, and test sets** using train\_test\_split from Scikit-Learn.

* **Training set**: 80% of the data
* **Validation set**: 10% of the data
* **Test set**: 10% of the data

The splits were organized in YOLO format with train.txt, val.txt, and test.txt containing absolute paths to the images.

**Model Used**

We used the **YOLOv8s object detection model** (Ultralytics).

* **Pretrained weights**: COCO pretrained weights were used as initialization.
* **Input size**: 640×640 pixels (default for YOLOv8s).

**Training Configuration**

* **Epochs**: 100
* **Batch size**:16
* **Optimizer**: Stochastic Gradient Descent
* **Learning rate scheduling**: handled internally by YOLO.
* **Loss functions**: YOLOv8s uses a combination of localization, classification, and confidence loss.

**Post-Processing**

YOLOv8 inherently applies **Non-Maximum Suppression (NMS)** to filter overlapping bounding boxes and retain the most confident predictions. Additional post-processing was applied to adapt predictions for dental radiographs:

* **Quadrant Division**: The panoramic X-ray was divided into **four quadrants** (upper-left, upper-right, lower-left, lower-right) based on midlines along the X and Y axes. This ensured that detected teeth were mapped to the correct region of the jaw.
* **IDF (Index Distribution Factor) Values**: To maintain proper tooth numbering, IDF-based positional checks were performed. Each detection’s relative position in the quadrant determined its assigned tooth ID, reducing misclassifications between neighbouring teeth.
* **Skipping Gaps for Missing Teeth**: If certain teeth were absent, the pipeline intelligently skipped numbering gaps instead of forcing continuous numbering. This ensured the predicted tooth sequence matched clinical reality.

**Evaluation**

The trained model was evaluated using the validation set. The following metrics were obtained:

* **Precision**: 88.5%
* **Recall**: 89.2%
* **mAP@50**: 91.6%
* **mAP@50–95**: 62.1%

## 4. Sample Result Images

## 

## 5. Training Curves

## 