Enhancing the Early Detection of Dental problem through Transfer Learning Techniques in Dental Radiography

MSc Research Project
Data Analytics

Himani Sharma
Student ID: X22224815

School of Computing
National College of Ireland

Supervisor: Mr. Vladimir Milosavljevic

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Himani Sharma		
Student ID:	X22224815		
Programme:	MSC-Data Analytics	**Year:**	2023-24
Module:	MSc. Research Project		
Supervisor:	Mr. Vladimir Milosavljevic		
Submission Due Date:			
Project Title:	Enhancing the Early Detection of Dental problem through Transfer Learning Techniques in Dental Radiography		
Word Count:	7700	**Page Count**	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Himani Sharma
Date:	11 August 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only

Signature:	
Date:	
Penalty Applied (if applicable):	

Enhancing the Early Detection of Dental Problems through Transfer Learning Techniques in Dental Radiography

Himani Sharma
ID: 22224815

Abstract

The poorest showing is represented by the sphere of public health with its focus on the most vital human health procedures and phenomena one's teeth. Such measurements allow dental panoramic radiography to be widely accepted among dentists in diagnosing and researching such diseases and with general exposure of the whole oral area use protective measures of low radiation dose and radiation time.To detect these matters, dentists have varieties of radiography like the panoramic views which do not take much time and have a low degree of radiation and perhaps, provide the visualize of the whole area in the mouth. It may take several hours and the whole process may tend to be tiresome especially when the veterinarian is interpreting the radiographs. Modern trends have allowed the dentists to complete the analyses in a shorter time with the support of various forms of Artificial Intelligence. It is rather sad today to such a group of people who require these services to part with an enormous amount of money to be treated by a doctor or maybe even get an x-ray. Large movies are when the malformations cannot be touched with the help of fingers In such a way, big movies are utilized by doctors.It assists in the management of the modern diseases such as caries, deep caries, impacted teeth and periapical lesions; hence development of cheap dental health care services.

Keywords- Caries Detection, Impacted Teeth, Periapical Lesions, Affordable Dental Healthcare.

1 Introduction

Pharmacological activities differ with the speed at which technology and research arena is expanding. It has become much easier to diagnose the illness and finding more practical solutions to it and this is due to Empowering Artificial intelligence that has been brought to lie between 75-86% by Ana Rita Pedro and Michelle B. Dias et al.(2023) in the field of medicine. Such inventions are even making the lives of humans much convenient and also putting healthcare in a better and more accessible place for everyone, thus fulfilling effective treatment and healthier lives. While the field of medicine operates at 0.2% to 0.1% error margin when it comes to technology in research it becomes a limitation to development and improvement. Likewise, the research has made some other efficient transformations in the dental sectors also it seems.Oral health conditions, such as dental caries and its sequelae, are common ailments, aggravated by conditions such as poverty or unsanitary habits, yet only 4.6% of global medical spending is anticipated to go towards oral healthcare by Abu Tareq, M.I., Islam, M.S. et al. (2024).

Yearly, a significant part of the world's population experiences some form of dental issues. According to the WHO Global Oral Health Status Report of 2022, it was calculated that about 2.5 billion populations globally, and most of them; seventy-five percent of those affected live in the middle-income nations. In this world more than 2 billion people have caries of permanent teeth and 514 million school-going children have caries of primary teeth. Education, preventive dental care and diets, and socioeconomic status affect the incidence of these conditions. This is why AI is critical in the identification of dental problems. From the above we can see how it assists in early diagnosis this means that even in dentistry complications are noted early and handled to avoid escalating. It also means that the ailments are diagnosed at an early stage, causing fast and efficient management. AI also helps in the provision of specialized dental care increasing the availability of specialized dentists where there are shortages. This is especially useful in parts of the globe where there is poor penetration by oral health facilities. In conclusion, AI acts as a tool that improves the quality and availability of dental care as well as contributes to the reduction of the costs. Peculiarly, costs on examining dental treatments are just steep bearing in mind that the only task being undertaken at such a cost is examination.

Panoramic x-ray devices are used to discover tooth defects. Efficiency is considered in relation to the use of artificial intelligence analysis, which is faster than the interpretation of X-ray films. The former is conversely used for the research of AI being algorithms for differentiating the normal and the abnormal teeth with the accuracy of nearly 98-99%. Thus, not only the diagnostics of teeth but also the treatment is facilitated. Earlier, in order to enable the computer systems that do not contain programming, the concept of machine intelligence incorporated new techniques such as Support Vector Machine (SVM) and statistical models to teach the computer and upgrade its learning and training capability. Other research done recently also used technologies such as Mask R-CNN for instance segmentation which was suitable for the identification and numbering of the teeth in panoramic x-rays. These object detection architectures find the position of the image, differentiating it from image classification where the only thing ascertained is the existence of object in image along with the features of the object but without giving details regarding the position of that object in the image. Among the image detection architectures, three of them are most widely used namely regional CNN, faster RCNN and You Only Look Once (YOLO). Food items and any obstruction in the mouth enhance the clarity of dental images in difficult cases. Accuracy testing can be considered with the help of precision score, the recall score, and the F1-Confidence Curve, (MAP) values of the model.

The research will therefore be based on an ability to screen for the dental problem categorized into 4 classes; Caries, Deep Caries, impacted and Periapical lesion where Fig.1 depicts impacted set of teeth. Dental caries is the most common disease of this system and oral cavity. Whether it is reversible or irreversible, pulpitis condition can develop from dental caries. Fig 2, Carries involving Loss of the diseased and decaying enamel and dentine is done by a high-speed dental drill and filled with amalgam, GIC (Glass Ionomer Cements) or composite resin. Impacted inform the region of teeth or enamel/ cavity in which the problem is experienced by M.Bouchahma, S. Hammouda.Periapical et al(2019). In Fig 3, Lesions are a containment against which microorganisms cannot infiltrate into the tissues; microorganisms cause the PA lesions, initially, or later by Kasra KARAMIFAR, Afsoon TONDARI et al.(2020).

2 Research Question and Objective

How transfer learning techniques in dental radiography help in the early detection of dental problems?

2.1 Objective

The ambition of this research study is the application of transfer learning methods in dental radiography for early diagnosis of dental issues. It deals with the creation and comparison of improved AI algorithms, including YOLO V8 and Faster R-CNN for differential diagnosis and Rt-DeTR for mapping the entire body for the identification and classification of dental disorders from panoramic X-rays. Focusing on caries, deep caries, impacted teeth, and periapical lesions, the models' accuracy is assessed in terms of precision, recall, F1 score, and (MAP) value along with confusion matrix. Its purpose is the enhancement of diagnostics and, thus, the advancement of the treatment process to become faster and more precise, finally improving the quality of the dental services. Furthermore, this study aims at improving the usage of specialized dental diagnostic tools focusing on such fields in the identified regions.

2.2 Motivation

The rationale for this research is that oral conditions, including caries and periodontal diseases are prevalent globally, for instance, dental caries affects 60-90% of school-going children and most adults as described by WHO Global Oral Health Status Report (2022). This can result in improved health care to those needy groups in the society. Prevention or care of these disorders in their elementary stages is very important to avoid worsening hence being complex to handle.

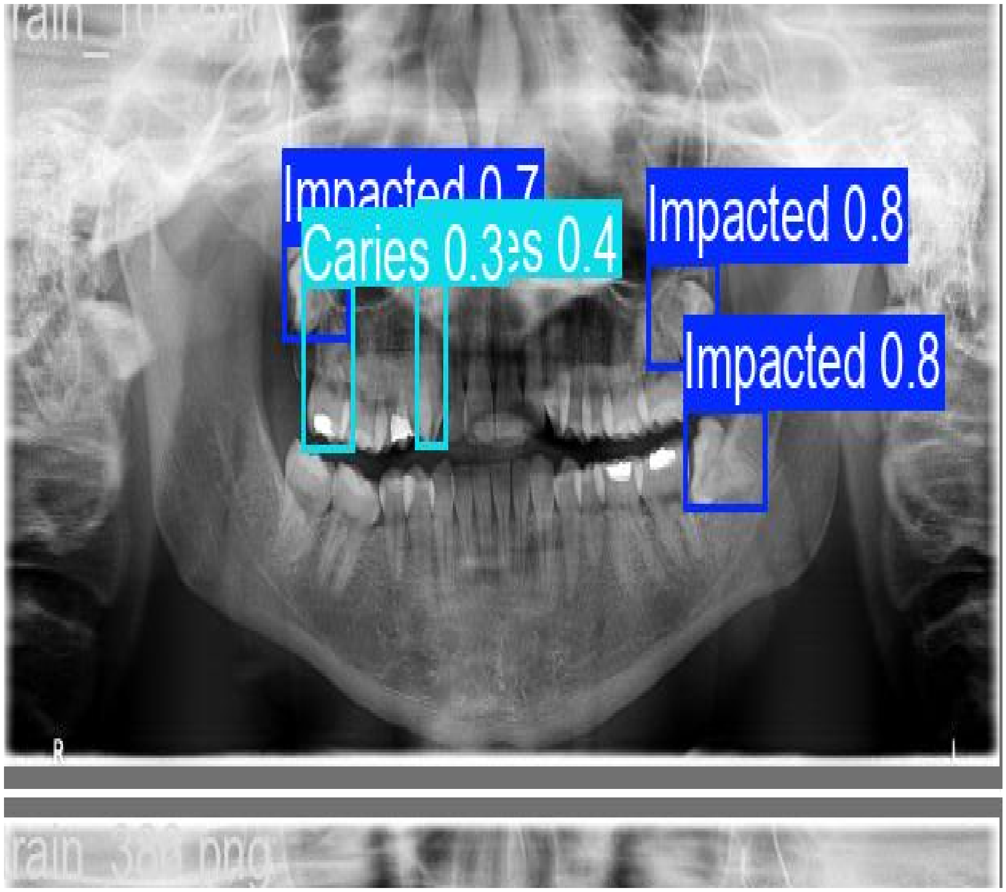

Figure 1: Impacted

3 Related Work

3.1 Traditional Techniques

In this paper, Shafi, K. et al., (2023) have explained how the use of image processing techniques for dental applications is rapidly increasing, more especially for the X-ray, CBCT and MRI databases. These imaging techniques offer enhanced, easy visualisation of the interior structures of oral cavity, which helps in diagnosing the conditions such as caries and periodontal diseases. The paper the authors used in their study is a systematic review of image-processing techniques concerning the PRISMA framework, of which CNNs for identifying tooth cracks are discussed. This means that MRI with the help of neural networks provided 95

Rahulsinh, Joshi, and Ahuja (2023) performed a scoping literature review to assess the outcomes of digital dental diagnosis and also an overview of the existing studies without imposing any restrictions on the type of study and publication year. They emphasize that PICO supports the formation of research questions and approaches and that image processing techniques, as a rule, are characterized by low risks in terms of the data flow and time. However, justifiably biasness in studies that involve human beings is an issue.

Karaoglu et al. (2022) prove that there is a necessity to advance the capabilities of sensing and detecting systems related to dental conditions. They investigate the applicability of the AI methodologies for the enhancement of the analysis of panoramic radiographs involving 2,702 scans. The authors of the study state that the use of Mask R-CNN along with heuristic algorithms proved more effective than baseline categories and increases the score accuracies.

In their study, Abramson et al. describe the anatomical principles and diagnostic use of dentistry and state that radiographic characteristics of milk and permanent teeth should be discussed. They stress on the fact of great attention should be paid to congenital dental abnormalities, including missing and supernumerary teeth and the DIC used in diagnosis.

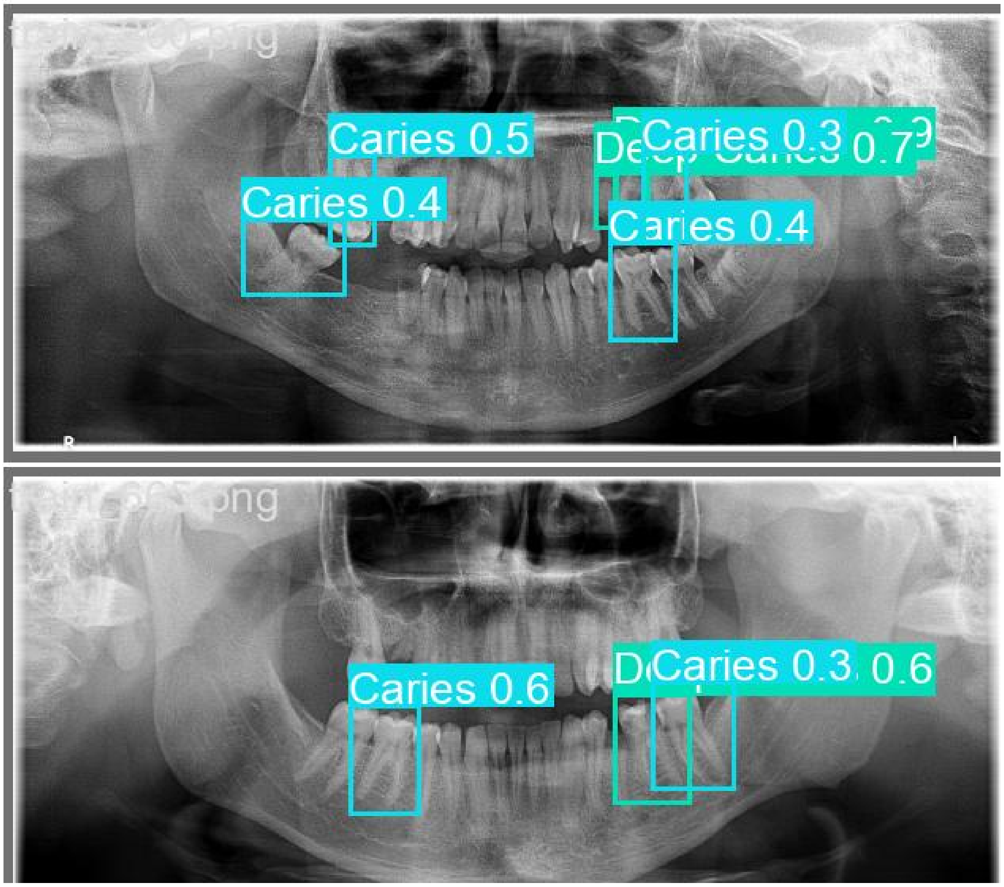

Figure 2: Caries and Deep Caries

You and Hao (2020) describe a paper in which a Convolutional Neural Network classifier was trained via transfer learning using a dataset of dental plaque areas. Their approach proved that AI could well show where dental plaque was, and the project presented a possible solution as to how parents could check on their children's dental hygiene by using a relatively affordable, portable intraoral camera.

3.2 Advanced Technologies

Huang et al. (2023) discuss the developments of the application of deep learning applying CNN, RNN, and GAN on dental image segmentation and recognition. Thus, they compare deep learning models with manual approaches stating that the former provide better results and work faster in the field of dental diagnostics. Ali et al. (2023) have considered the employment of YOLOv7-t and YOLOv7-f for the detection and numbering of teeth from dental panoramic images. Using their model, based on the dataset of 3,138 radiographs, the authors obtained the high precision and F1-scores, thus, highlighting the application potential of AI in the field of dental X-ray analysis.

Dental caries is discussed by Latke et al. (2023) and is strongly associated with root canal procedures. Their work integrates the two classifiers namely, SVM and ANN yielding 95

Qayyum et al. (2023) tackle the difficulty of finding posterior proximal caries diagnosis through a Dental Caries Detection Dataset (DCD2). They compare and identify the current novel deep learning architectures such as PCA, MLP, U-net which provide a strategic approach towards caries detection. Therefore, the study shows the importance of the presence of numerous and various samples to enhance the diagnostic detection.

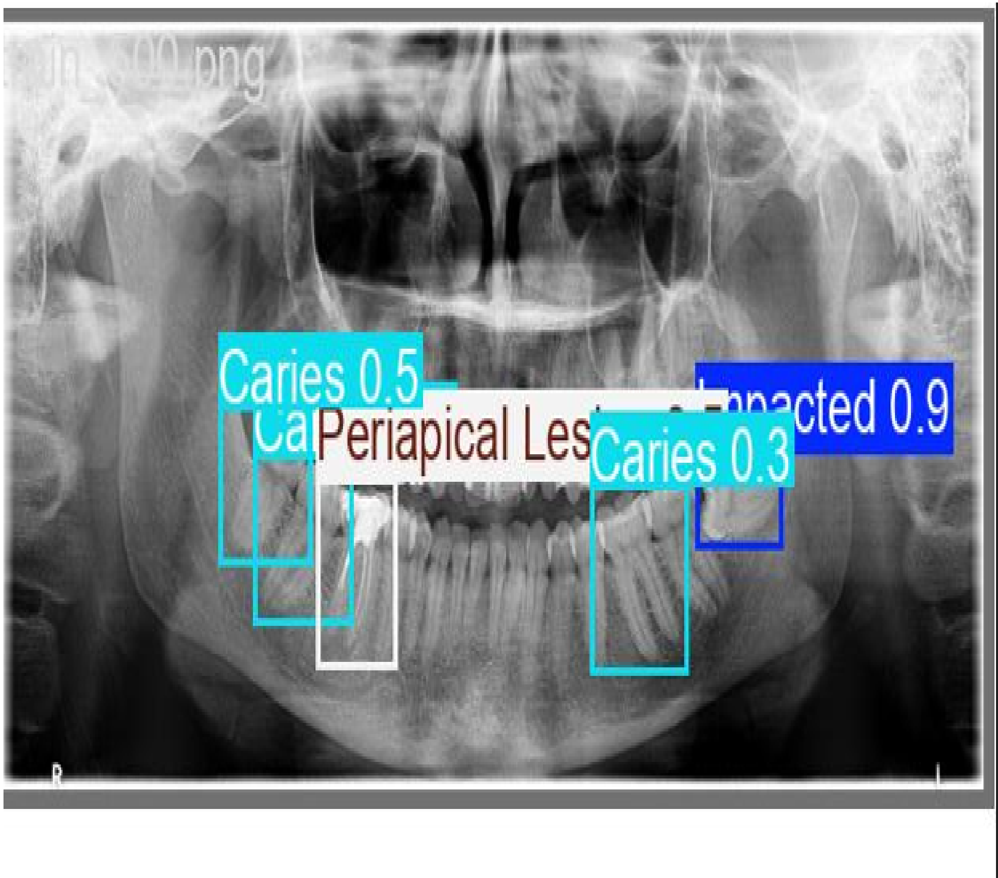

Figure 3: Peripherial Lesion

Model Type	Description and Evaluation
One-Stage Models	YOLO, RetinaNet The emphasis is on object detection but, at the same time, preserving accuracy while increasing speed
Experimental Model	RetinaNet Suitable for handling large datasets efficiently
Two-Stage Models	RNN-based models Achievable by re-estimating it for one-stage models with pre-trained data
Two-Stage Modelling	Focuses on achieving higher detection accuracy Make updated trained data from the one-stage model to minimize error rate

Table 1: Evaluation of One-Stage and Two-Stage Models

3.3 GAP Analysis

Modern research mainly employs the single- or two-stage same-stage models for speed at the cost of efficiency such as YOLO or Retina-Net, or the two-stage models for greater accuracy. This work attempts to include both strategies, Employ Retina-Net for huge datasets while employing RNN-type models for accuracy. The basic principle adopted in the two-stage modeling is that detection precision will be prioritized over the one-stage model's preparatory training, which will be pre-trained on given data. To assess the findings of the study the following measures shall be used as laid down in the research section.

4 Research Methodology

4.1 Methodology

In Fig. 4, herein this research, the methodology used is CRISP-DM. Figure 3 offers a depiction on the research framework and process regarding feature extraction, data acquisition, and analysis, and modelling. This research work consists a total of five (5) parts and each part is relevant and crucial to this project.

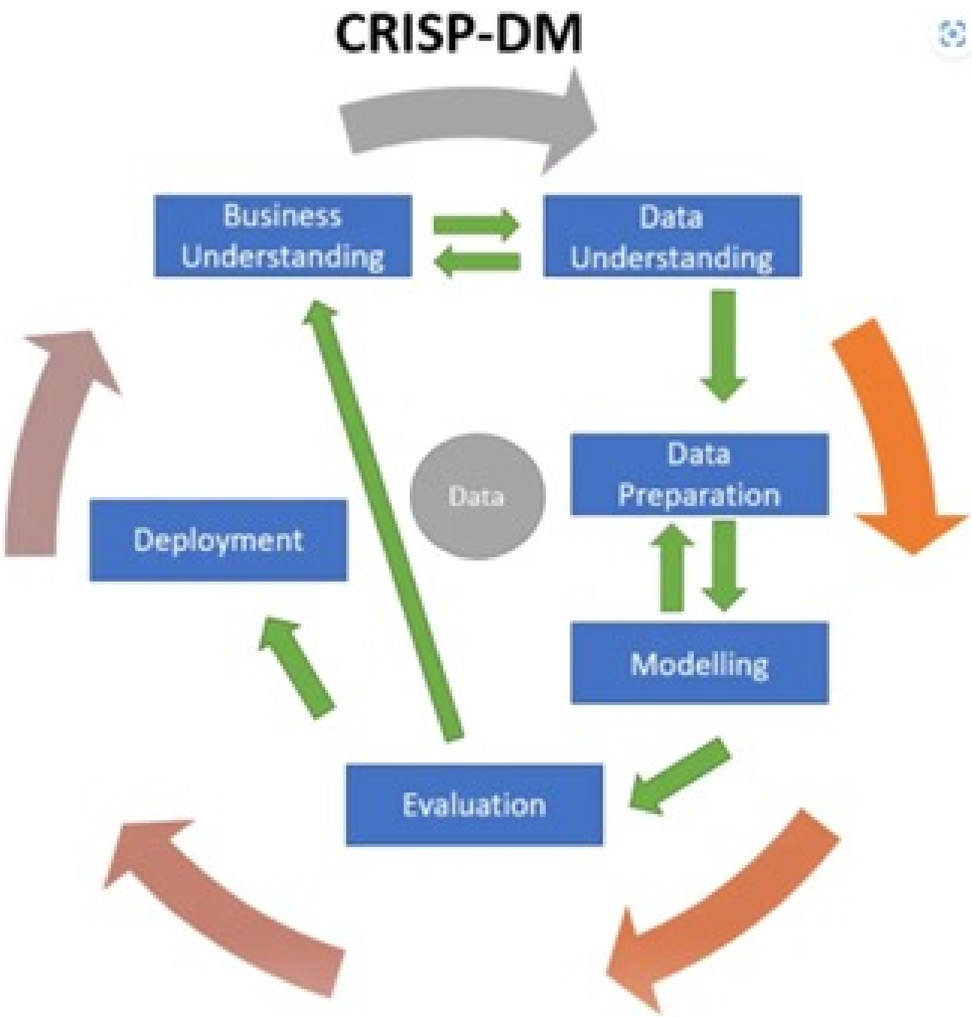

Figure 4: CRISP-DM Methodology

4.1.1 Data Gathering

The first operation of this project is archiving the Dentex dataset which in Fig. 5 is sourced from Kaggle an open platform. This dataset consists of a total of thirty five hundred panoramic X-ray images for diagnosis of several dental related diseases. The images are categorized into four classes: smear samples were obtained; 500 samples from impacted teeth, 1750 from caries, 750 from lesions and 250 from deep caries. Yet, 1571 images did not have labels for them and were excluded in order to match the objectives of the supervised learning dataset. Pertinently, it is necessary to ensure that the collected dataset does not violate Kaggle's licensing policies and laws.

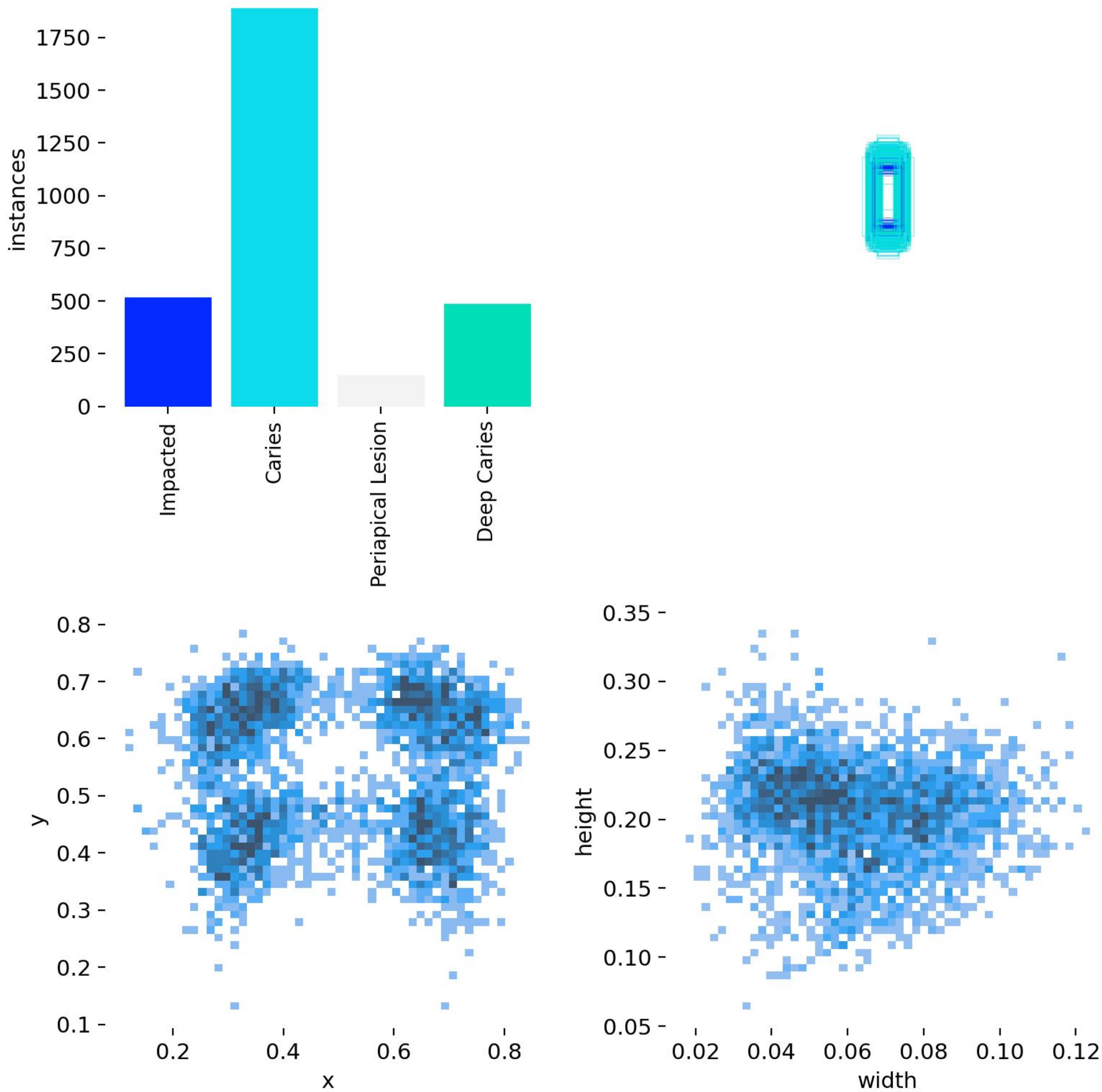

Figure 5: Dental Data for Various Conditions

4.1.2 Data Pre-Processing

Data pre-processing is a crucial phase in the pipeline that involves preparing raw data for analysis to ensure that it is clean, accurate, and suitable for training models. This process is vital because the quality of the data directly affects the performance and reliability of machine learning algorithms. Two specific functions that play a significant role in this pre-processing stage are `remove_images_without_annotations()` and `sample_and_copy_images_labels()`, each of which addresses different aspects of data quality and usability.

4.1.3 Data Pre-Processing

Data pre-processing is a step in the data pipeline that aims at cleaning the data for training models since it is the first step towards the modeling pipeline. This process is rather important because the quality of data determines the efficacy and effectiveness of machine learning algorithms. There are two specific functions that bear a great deal of responsibility in this pre-processing phase, namely `remove_images_without_annotations()` and `sample_and_copy_images_labels()`, each of which are trying to resolve different issues pertaining to the suitability of the data.

Null Values Regarding the problems in Fig. 6, Null values in datasets, particularly image data, can create huge issues. For instance, missing annotation file can cause issues when the total quantities of images and their corresponding annotation files do not match. To address this problem, the function `remove_images_without_annotations()` is created to search for such images and erase them in case they do not contain requisite labels. This helps to avoid having mismatched data that could compromise the performance levels of the model intended to work on the given dataset.

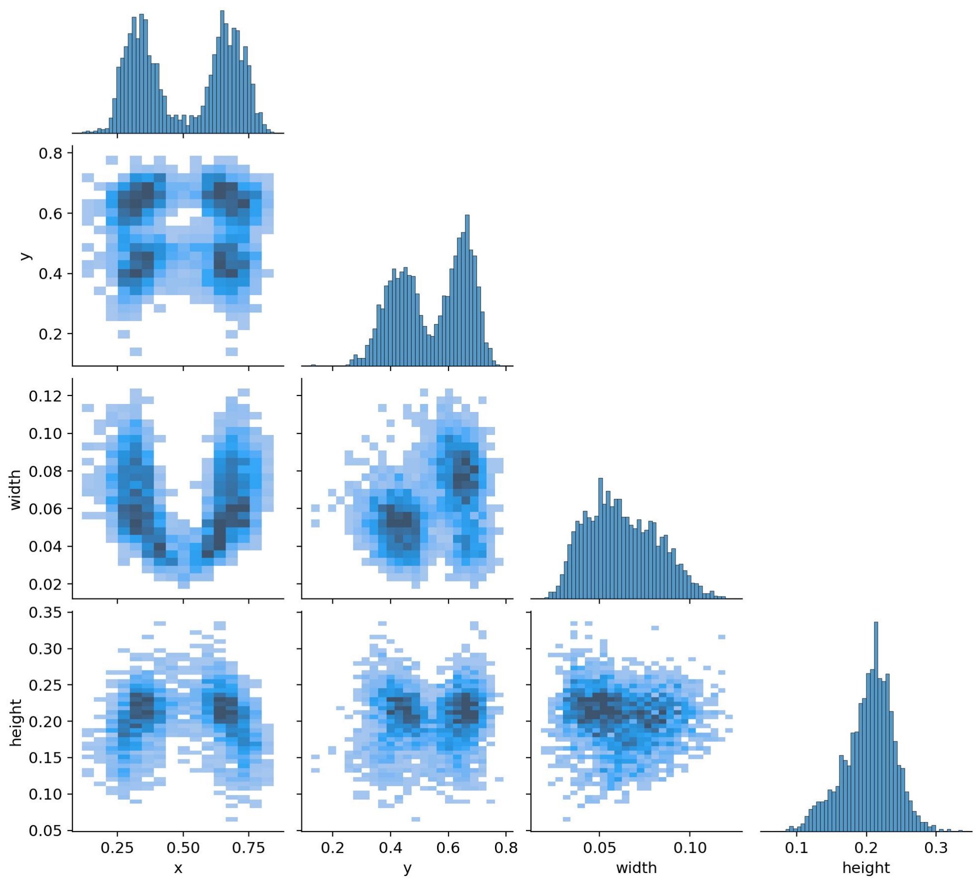

Figure 6: Unlabeled Dental Images

Augmentation Data augmentation is a method used to increase the size and diversity of a training dataset without actually collecting new data. By applying transformations like rotations, flips, and shifts to existing images. In Fig.7, we can create many variations of the same image. This helps the model learn to recognize objects in different scenarios and reduces the risk of overfitting to the training data. The `augment_images()` function contributes to data modelling by generating augmented versions of images.

Splitting Data

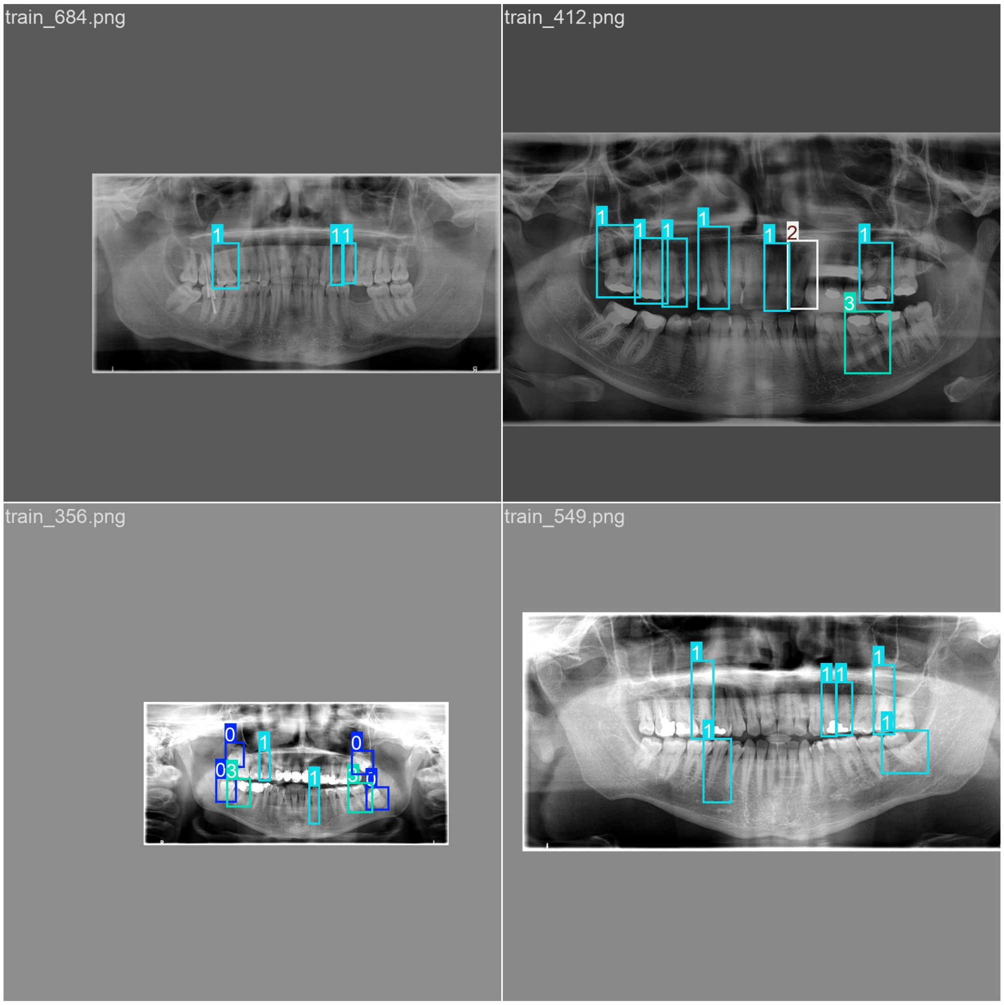

Figure 7: Augmentated X-Rays

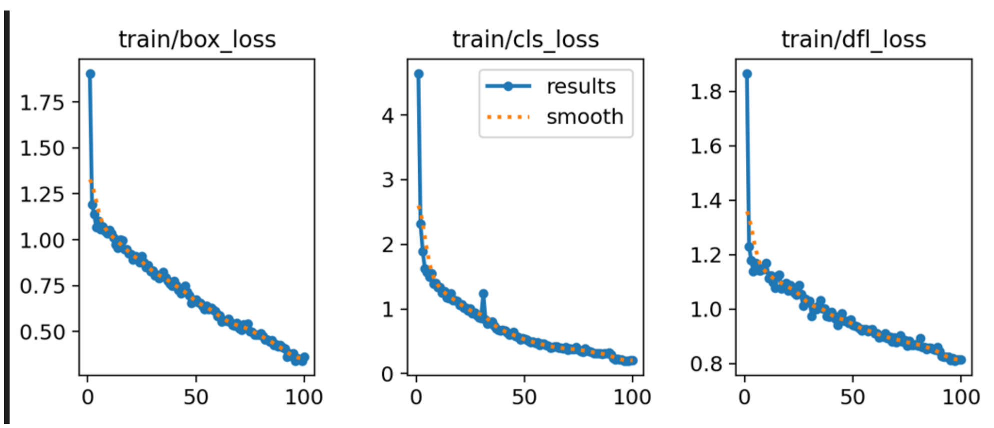

Figure 8: Dental Train Data for Various Conditions

Augmentation In data augmentation the size and/or the variety of the training dataset is increased by employing some methods which do not require the collection of new data. Through the use of rotations, flips and shifts on other images. From Fig. 7 below, numerous modification can be made to obtain the same basic image which means that there are a number of choices for a given picture. This assists the model to learn on how objects are positioned differently and also prevents overemphasizing on the training phase data. The `augment_images()` function also benefits data modelling since it is responsible for producing the augmented images.

As seen in Fig. 8, division of datasets into training set and validation set is very crucial when training a model. It is also customary to use 85 or 75% of the images for training and the remaining 15% for validation; concerning the help them, the function `sample_and_copy_images_labels()` randomly takes 15% of images from the training set and then copies these images along with their labels to the validation image folder. This makes certain that the validation set is in fact diverse with regards to the other data existing.

4.1.4 Data Transformation

Data transformation is the last but a very important step before feeding the data into a model; it involves transforming the data into forms that are acceptable by different models or tasks. It is at this point where the `process()` function assumes a very important function. In Fig. 9, It converts the annotations given in the COCO format to YOLO format. YOLO stands for You Only Look Once and is a model format for object detection, for which the annotations must adhere to a particular format.

Another crucial change is done by `convert_yolo_to_pascal_voc()` function Here the annotations of the objects are changed according to the Pascal VOC format. This function takes annotations in the YOLO format and translates it into another format often used in computer vision tasks known as Pascal VOC format.

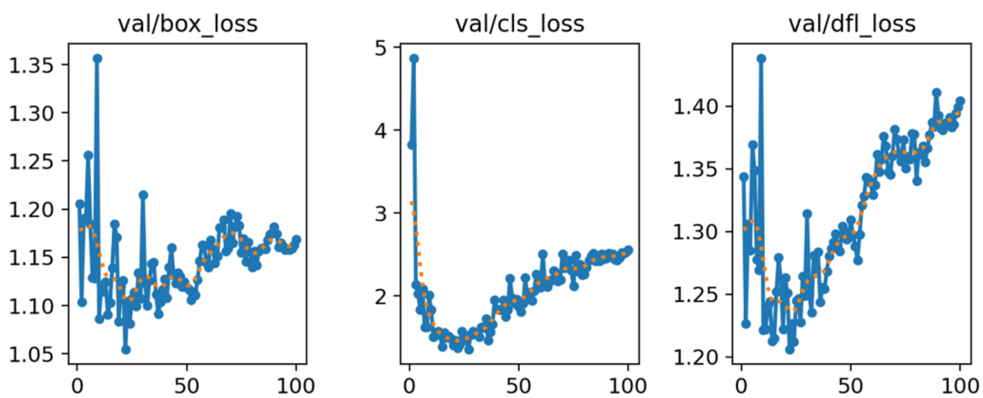

Figure 9: Data Transformation-After Validation

4.1.5 Data Modelling

Data Modelling is set where the prepared data would undergo a set of model trainings. This research paper is using 3 models which are YOLOV8, RT-DETR and Faster-RCNN.

YOLOv8 Another model given by YOLOv8 is the semantic segmentation model, the YOLOv8-Seg model. The backbone instead is a CSPDarknet53 feature extractor and the neck is a modified C2f instead of the YOLO one. The C2f is succeeded by two segmentation heads responsible for predicting the semantic segmentations masks of the input image. Like YOLOv8, it has the same number of detection heads consisting of five in total while one is a predicting layer by Terven, J. , Córdova-Esparza et al. (2023) YOLO is currently among the best model architectures for object detection and other related challenges. It has a neural network structure to achieve high performance and provide overall high speed of work. The training and evaluation processes are performed on three versions of YOLOv8 for object detection – small (v8s), medium (v8m), and large (v8l) – on the YOLOv8-specific custom dataset mentioned in the `custom. yaml` file.

Real-Time Detection Transformer (RT-DETR) RT-DETR outperforms not only YOLO detectors of similar size, but also other detectors that have fewer parameters. However, Transformer-based end-to-end object detection models still have extraction difficulties as well as their high computational cost and the problem of insufficient accuracy in detecting small targets by Wang, S. , Jiang et al. (2024). RT-DETR is a new and powerful end-to-end object detector with regard to real-time performance accompanied by high accuracy. It is based on the idea of DETR, using the conv-based backbone, and the efficient hybrid encoder gains the real-time speed. The training process can be set up to run for a number of iterations, for the purpose of the upcoming experiments it will be set to 100 epochs, the images are rescaled to be 1280 pixels in width and at a fixed rate of 16.

Faster R-CNN Faster R-CNN is another deep convolutional network and is perceived be the user as a single unified end-to-end network. The process of training consists of data augmentation, several batches of data for per epoch, and they adopted SGD or AdamW optimizers with cosine annealing as the learning rate sheduler. The regularly utilized type of neural networks encompass artificial neural network (ANN), convolutional neural network (CNN) and recurrent neural networks by Imran Shafi 1 and Anum Fatima et al (2022).

4.2 Evaluation

Evaluation The performance of each model is checked after training through a validation phase, where one of the important parameters of the model is computed: the mean Average Precision (mAP) over all the Intersection over Union (IoU) thresholds ranging from 0.5 to 0.95. These include mAP at 50 IoU (mAP_{50}), at 70 IoU (mAP_{75}), and over all IoU thresholds ($\text{mAP}_{50\text{-}95}$). Before Feature Pyramid Networks (FPN), a majority of deep learning-based detectors performed detection only on specific features of an object, such as a vehicle, rather than on the entire object itself. The main layer profiles of the networks, as described by Zhengxia Zou, Keyan Chen et al. (2023), show that Mean Average Precision (mAP) is an extensive measure used to evaluate the performance of object detection models. It measures the model's capacity to accurately identify and localize objects in the defined classes, and is calculated for all classes jointly across various IoU thresholds.

(a) Average Precision (AP):

- **Precision:** Precision indicates how many of the detected objects are correct, which is why this metric is referred to as the degree of object detection. It is defined as the proportion of true positive detections out of all detections, including both true and false positives.

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

- **Recall:** Recall measures the extent to which the real objects have been detected. It is defined as the ratio of true positive detections to the sum of true positives and false negatives.

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

- **Average Precision (AP):** For a given class, AP is the area under the precision-recall curve, obtained by varying the detection confidence threshold. It summarizes the accuracy of the model's precision-recall performance for that specific class.

This model also calculates the Mean Average Precision (mAP), which is the mean of the AP scores across all classes. Thus, it measures the overall performance of the model and provides a difference between the observed and expected values, intending to reflect performance across different object classes.

2. mAP Performance at Different IoU Thresholds: Intersection over Union (IoU) measures how accurate the bounding box predicted by the model is compared to the actual bounding box. It is defined as the ratio of the area of intersection between the predicted and ground truth boxes to the area of their union.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

- **mAP$_{50}$:** This stands for Mean Average Precision at an IoU of 50. It measures how well the model performs when the number of true positives is large, as well as when the number of true negatives is substantial. In this case, the average intersection of the ground truth and predicted bounding boxes is at least 50

- **(mAP)50**: The (mAP) at 50% IoU threshold. It measures how well the model performs when the overlap between the predicted and ground truth boxes is at least 50%. This is a more lenient criterion.
- **(mAP)50-95**: This is the mean Average Precision averaged across multiple IoU thresholds from 50% to 95%, typically in 5% increments (i.e., 50%, 55%, 60%, ..., 95%). It provides a more comprehensive evaluation of the model's performance across various levels of localization precision.
- **(mAP)75**: The (mAP) at a 75% IoU threshold. This metric is stricter and measures performance when the overlap between the predicted and ground truth boxes must be at least 75%. It reflects how well the model performs under stricter localization conditions.

5 Design Specification

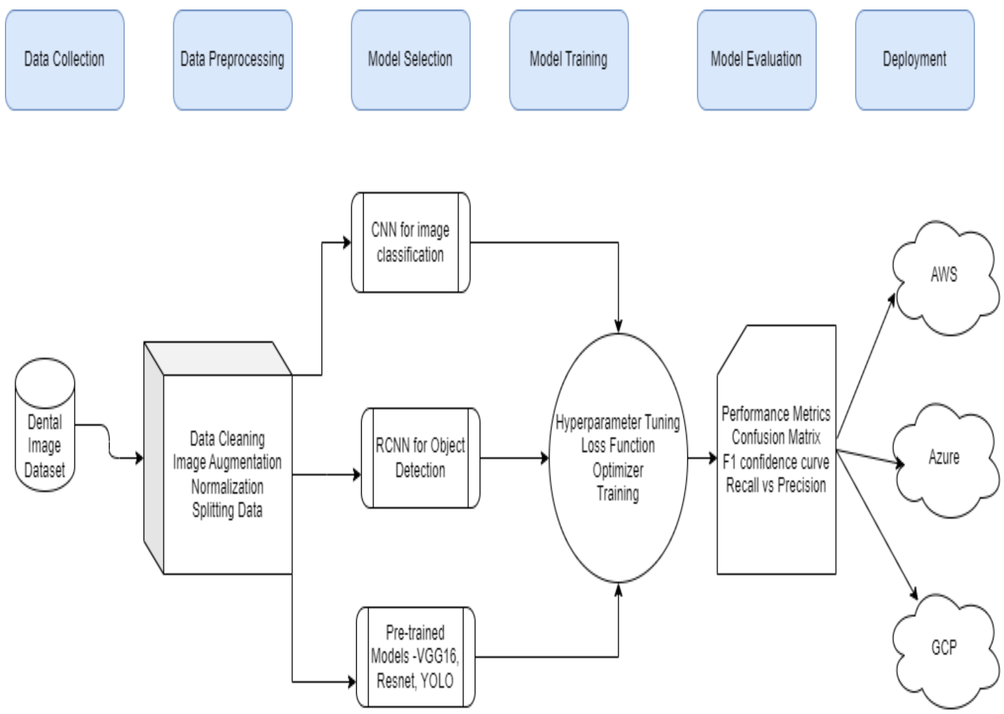

Figure 10: Design Specification

This paper describes the general idea of approach required for a classification project carrying out on dental X-ray images using a methodological framework known as CRISP-DM that follows the business structure of data mining gathering, pre-processing, transforming data, modeling, and assessing.

As depicted in Fig. 10, the first preprocessing step is to compile an image database of 3500 images to be classified into four classes of Oral cavity pathology images.Preprocessing steps include, scaling of pixel intensities and the enhancing the data in terms of robustness by applying some transformations to the image data and then use of the appropriate technique to split the image data in to training set and validation set. Subsequently, the image annotations are converted into formats that different models can process such as Labeled SHD or SHD.

We train three types of models: Here the setups of three models: For YOLOv8 to be used in TRAIN mode with a batch size of 64 and RT-DETR to be used in the same mode but the batch size is 64 Similarly, the model that is to be used in TRAIN mode is Faster R-CNN with a batch size of 16. Finally, we conclude by evaluating the model's ability to detect and segment the dental areas of concern based on evaluation measures such as mAP, precision, recall, and IoU. This is carried in Python using relevant machine learning libraries to ensure that the system under development is efficient, reliable, and user-friendly.

6 Implementation

6.1 Data Collection

The **Data Collection** requires the acquisition and structuring of the dataset known as Dentex which is freely accessible on Kaggle. This dataset contains 3,500 panoramic X-ray images categorized into four classes: It concerns the analysis of Impacted Teeth, Caries, Lesions, and Deep Caries. Each class is an instance of a particular kind of dental condition. For supervised learning any images that do not have a label assigned to them are dropped. This step important in order to increase the accuracy of the further data analysis and training of the model.

6.2 Data Pre-Processing

Removing Images without Annotations: The images that are not labeled are discarded to make sure that all the images which are used in training are labeled images. Values presume by digital image pixels referring brightness or intensity are normalized to be between 0 and 1. This standardization aids the Machine Learning algorithms to perform better. In order to decrease the probability of overfitting there are certain transformations included into the pipeline as follows : rotation, flipping, brightness adjustament. The dataset is split into 85 training data and 15 validation data so as to make a clear distinction between the training and test sets.

6.3 Data Transformation

Data Transformations phase involves the preprocessing of the data to make it compatible for various object detection models. Annotation data is converted to YOLO format whereby coordinates of the objects are in relation to the size of the image. Also, annotations that are realized with YOLO format saved as Pascal VOC format to provide possibility to use it with other models or instruments in the sphere of computer vision.

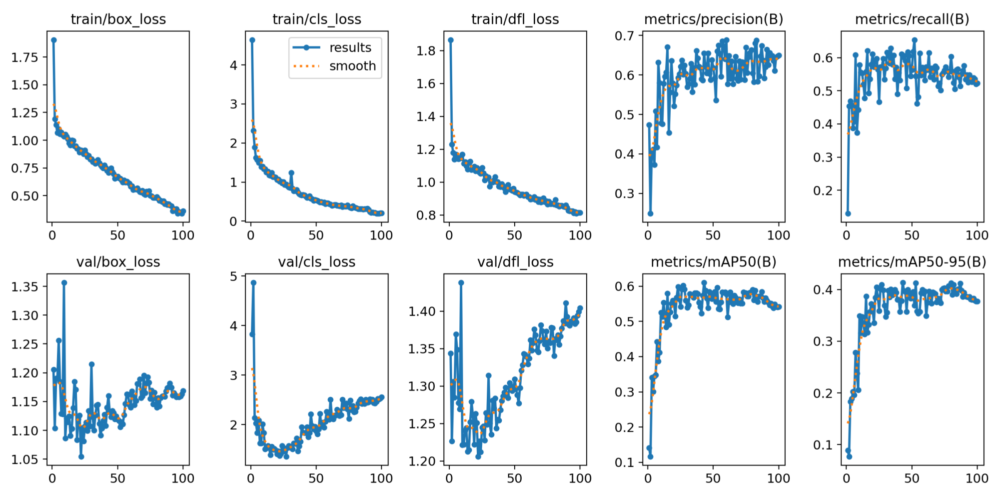

Figure 11: Evaluated Results

6.4 Data Modeling

The **Data Modeling** phase involves training the object detection models on the prepared dataset:

- **Training YOLOv8:** Each version, however, is an attempt between accuracy in evaluating the model and the time taken in assessing the inference time of the model. In training it is necessary to make use of pre-trained weights so as to start from the existing information. The process includes configuring hyperparameters such as learning rate and batch size, running the training loop, and applying early stopping based on validation performance. Hyperparameters that are used include; The process also entails tuning of such hyperparameters as learning rate and batch size, training loop, and early stopping on the basis of the validation performance:

- **YOLOv8s:** A small version of YOLOv8 that are optimized for a faster run time with less computation cost compared to YOLOv8 while reasonable accuracy.
- **YOLOv8m:** An intermediate, that is faster than the small variety yet reduces the time taken as compared to the large version while ideally increasing the accuracy of the calculations with reasonable computational complexity.
- **YOLOv8l:** It is a version of YOLOv8 with the highest accuracy, aimed at the best detection performance; at the same time, this version consumes more computational resources than the YOLOv8-tiny ones and has lower speed of inference.

  All versions are trained with pre-loaded weights, and the training process includes early stopping to prevent overfitting.
- **Training RT-DETR:** This model converges more quickly by continuing the weights of other nets achieved by PCG which is pre-trained. Training implies setting of hyperparameters, performance of the training process, and detection of overtraining with the help of the stop criteria.
- **Training Faster R-CNN:** Training enhances efficiency especially via the use of pre-trained weight. It involves setting hyperparameters,tuning them and carrying out the training loop as well as incorporation of early stopping in a bid to effect a check on overfitting.

All the models are trained under the consideration of achieving the best performances with the help of the pre-trained weights and early stopping. This approach provides for efficient training, high accuracy to avoid overfitting of the training data set as well as being fast. The measure of performance is in terms of mean Average Precision (mAP), precision, recall, and even Intersection over Union (IoU).

.

7 Evaluation

7.1 Experiment A

As shown in Fig. 11, This yields to an mAP@r for recall rates of 0.5646 and an mAP@50-95 of 0.39546. These were obtained with 80-20 ratio used for train-test data split, thus giving a balance to the evaluation though this exposed inherent weakness of the model majorly illustrated by the shortcomings it exhibited in correctly identifying the box pattern. The moderate performance indicated a possibility of the data needing a deeper model and better layering to enhance the amount of detected data. The "Impacted" class fared better than others, probably because of the scarce training data set; this shows that the model is more challenged by such classes as they may contain few or highly complex samples. The plot depicts that the Impacted class has high F1 score all the time with different thresholds which reveal that the performance of the system is good. However, the Caries, Periapical Lesion, and Deep Caries classes (Orange, Green, and Red curves) are indicated to be of lower and variable F1 scores.This is realized at a 0.95 confidence level and this is 0.64. Regarding the evaluation of the results, the cross validation score of 147 is presented as the most optimal, since it achieves a good balance between precision and recall.The best results identified from the analysis of F1 scores, reveal that the Caries class remains comparatively low and consistent over time, which indicates a progressive decrease; the Periapical Lesion shows fair to poor and oscillating performance; finally, the Deep Caries reveals a moderate F1 score which depicts slow and steady declining pattern.

It can be seen that for the YOLOv8 Large model, all variants achieve better results than other models with a precision of 0.63. The recall is 0.60, the number being 63182. It achieved an AP of 60,589, and an mAP50 of 0. For extension cord 61353 with mAP50-95 equals to 0.40. Trained with 85-15 data split, this model shows better predictive performance on all the measures, as derived from a more complex model architecture and increased ability to work with the data. Thus, its capability of moving through epoch cycles makes it productive in forecasting a box score: evidenced by the presented high value of mAP50 of 0.61. Nevertheless, the variance of the "Caries" class implies variability in the likelihood of detection, which

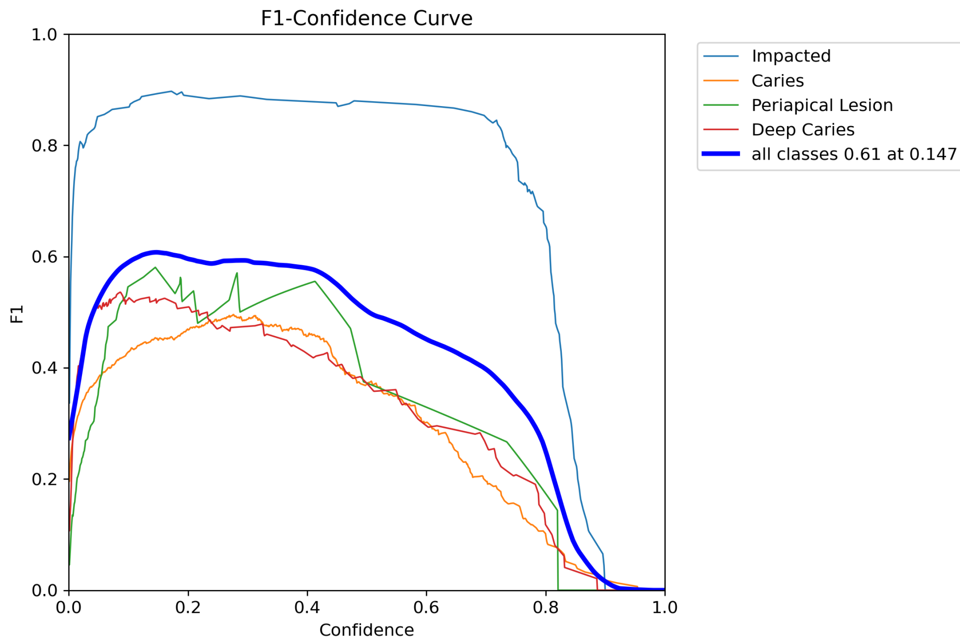

Figure 12: Confidence Curve - YOLOv8 (Small)

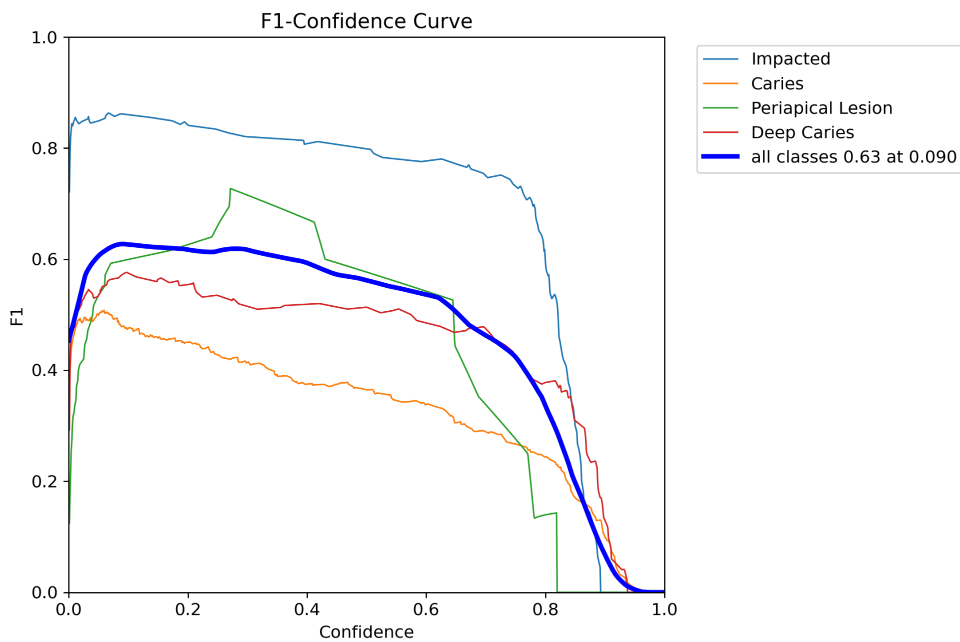

Figure 13: Confidence Curve - YOLOv8 (Medium)

can be attributed to the fact that the data set in this group is considerably bigger and less homogenous in structure. Caries, Periapical Lesion, and Deep Caries predictive classes' F1 value is comparatively lower and highly volatile. It can be noted, that the maximum average F1 score is received with the help of coefficient at the confidence level of 0.164 as pointing out the optimal cut-off point that results in the best trade-off between precision and the recall rate.

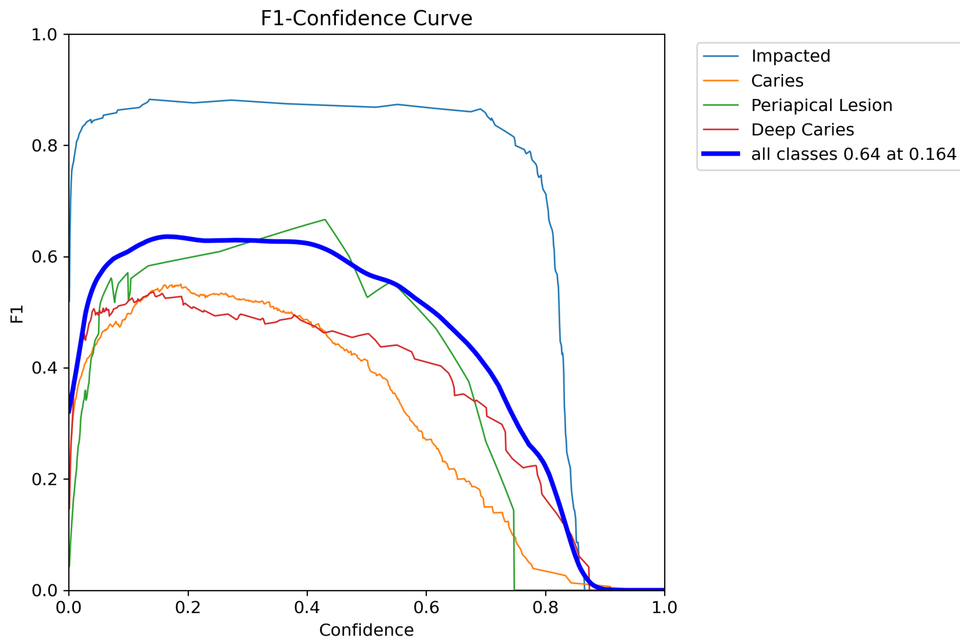

Figure 14: Confidence Curve - YOLOv8 (Large)

7.2 Experiment B

7.3 Experiment C

As for the models' precision, Faster R-CNN demonstrates the lowest value, equal to 0.491 and the carryforward number or recall 0. The results showed that the year 1847 was difficult in terms of being able to find and capture the true positive cases. Its mAP50 is 0.3561; the mAP50-95 score is not indicated, therefore it would appear that the proposed network has some difficulties in terms of flexibility in connection to a growing amount of intersection of identified and predicted boxes. Faster R-CNN has a traditional single-stage architecture that is trained with an 85-15 data split, and its efficiency is inferior to that of the two-stage system in this case, especially considering the network's inability to work adequately with large epoch cycles and small batch sizes. This leads to instability and unsustainability of its performances as compared to other models and that's why it has its demerit in this particular application. In Fig. 15, We show the mean Average Precision ((mAP)) of the model over 100 epochs of training using the Faster R-CNN model. The orange line refers to (mAP)50 showing the principle of precision at the IoU = 0. 5 to the blue one corresponding to (mAP)50-95, which signifies the average precision at the different IoU thresholds ranging from 0. 5 to 0.95. Comparison between (mAP)50 and (mAP)50-95 indicates that (mAP)50 has higher value and also has almost constant values as compared to values of (mAP)50-95 which have more fluctuations. This implies that the model is better suited for a models with lower IoU and has averate but relatively lower precision when combining across multiple logistic curves which are at higher IoU thresholds. In general, the mAP and mAP50-95 performance increases and levels off with iterations, although the latter metric fluctuates more on the strictly tighter curve. However,Incorporation of these architectures, fast or faster CNN reveals fewer convergence times, as per the characteristics of the fast or faster CNN, in comparison with RCNN by Tausif Diwan, G. Anirudh et al. (2022).

7.4 Discussion

Fig 16 exhibits the performance metrics graphs where it shows of defined model.

Faster RCNN: Shows the weakest performance with a significantly lower precision of 0.491 and recall (mAP) of 0.1847. The (mAP) 50 is 0.3561, and (mAP) 50-95 is not reported. Since the train-test split was

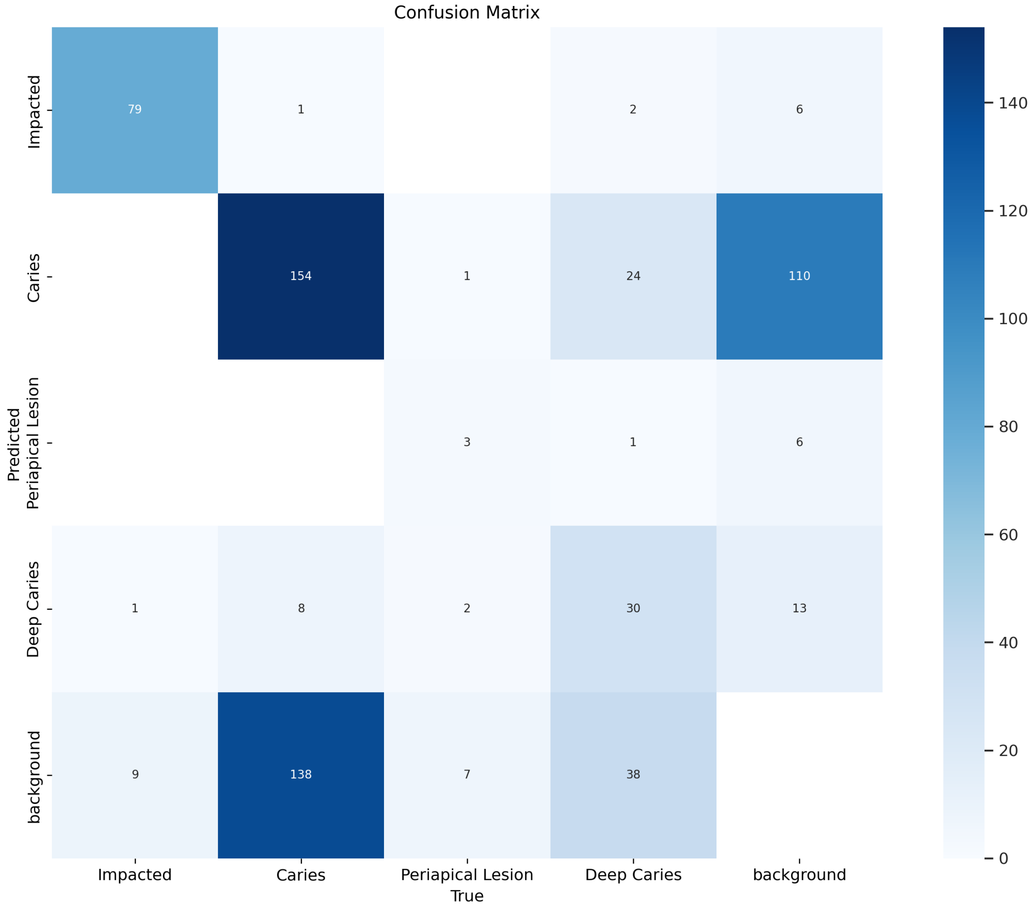

Figure 15: Confusion Matrix - RT-DETR

(85-15) Since the model was traditional and a conventional model with one stage architecture. Faster RCNN compared to other models lacks suitability thus data can not be run for larger epochs cycles holding less batch size. Thus, it lacks model stability and robust performance. However, Studies have demonstrated that CNN models can surpass the accuracy of manual methods in classifying chronological age based on dental images by Kokomoto, K., Kariya et al.(2024)

YOLO V8 Small: It Shows moderate performance with precision of 0.60248, recall of 0.54488, (mAP) 50 of 0.5646, and (mAP) 50-95 of 0.39546 since the train-test split was (80-20) which makes data moralized for evaluation pattern and shows that data is nearly needs better layered model to perform box pattern and impacted has outperformed due to fewer training dataset.

YOLO V8 Medium: It has the highest precision of 0.65002 but lower recall of 0.52261. The (mAP) 50 is 0.54147 and (mAP) 50-95 is 0.37731. 39546 since the train-test split was (85-15) . YOLO V8 Medium as compared to YOLO V8 small is better with layered structure architecture and can accumulate better data driven research by enhancing the augmentation feature thus given better precision value but lower recall since it was not able to detect a few dental images box scores and likely it requires more training data to outperform.

YOLO V8 Large: Outperforms other models with a precision of 0.63182, recall of 0.60589, (mAP) 50 of 0.61353, and (mAP) 50-95 of 0.40979, making it the best overall. 39546 since the train-test split was (85-15) where it is one of best model in terms of all model predictions with best architecture and enhanced data

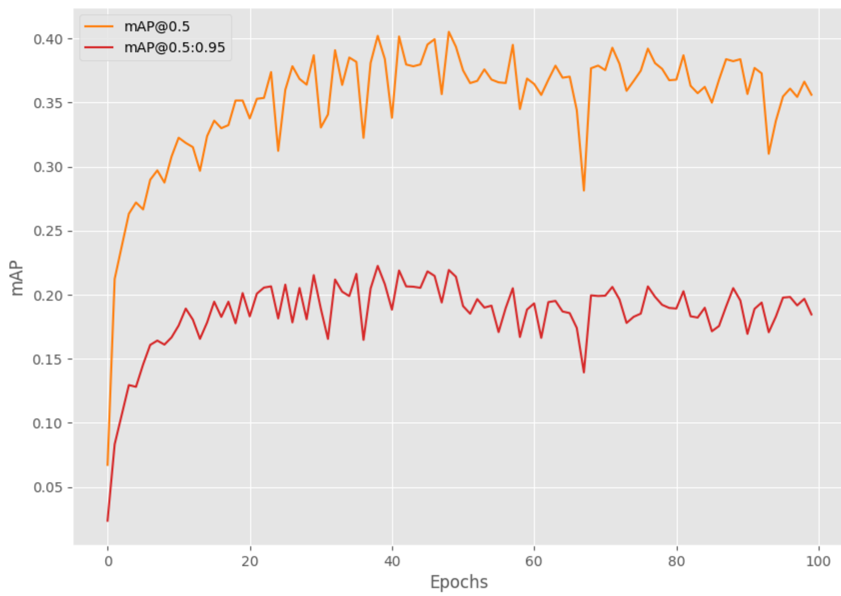

Figure 16: (mAP) Graph - Faster R-CNN Model

capacity. It has better rate of running the epochs cycle due to its versality and Box score prediction since the mAP value lands at 0.61 and Caries shows variable pattern due to its larger data presence.

RT-DETR: Achieves good balance with a precision of 0.61839, recall of 0.57952, (mAP) 50 of 0.59149, and (mAP) 50-95 of 0.38697. since the train-test split was (80-20) where its performance is be compared to YOLO V8 where recall rate is highest which means it detect true better box images though it lacks mAP value where it lacks augmentation feature, but overall performance is good.

8 Conclusion

The differences in the object detection models raise questions of how efficients various models are at detecting dental conditions. Cozycov Large YOLO is the best overall with good precision and recall of 0.63182 and 0.60589 respectively. The high figures of mAP are demonstrated, in particular, 0.61353 for mAP@50 and 0.40979 for mAP@50-95 which points to the fact that it is efficient in dealing with large amounts of data and processing complicated images. Nevertheless, it gave slightly less accurate results in the Caries identification, pointing to the variations in the data set used for training. YOLOv8 Medium performs exceptionally well in precision rating with a score of 0.65002 This Specificity simply means that it is very accurate whenever it makes the identification of the flaw. However, there is lower numbers for recall that is 0.52261; which tells that the algorithm may fail to detect at times. This model is good if accuracy is given preference over recall because, though it is very accurate in results, it requires more training data to enhance recall.At last, Faster R-CNN indicates the least precision as 0.491 and a recall of 0.1847. It is relatively slow especially with large data and when tested for longer training times, perhaps due to its inefficiency in handling the more complicated dental images. Last but not the least, YOLOv8 Large model along with YOLOv8 Medium are the most effective models out of the four ones enlisted here, but it is also important to realize that each of these models could perform even better if it has access to better datasets and if more sophisticated methods of training can be applied. Faster R-CNN, however, has many shortcomings and possibly is not an ideal solution for this problem.

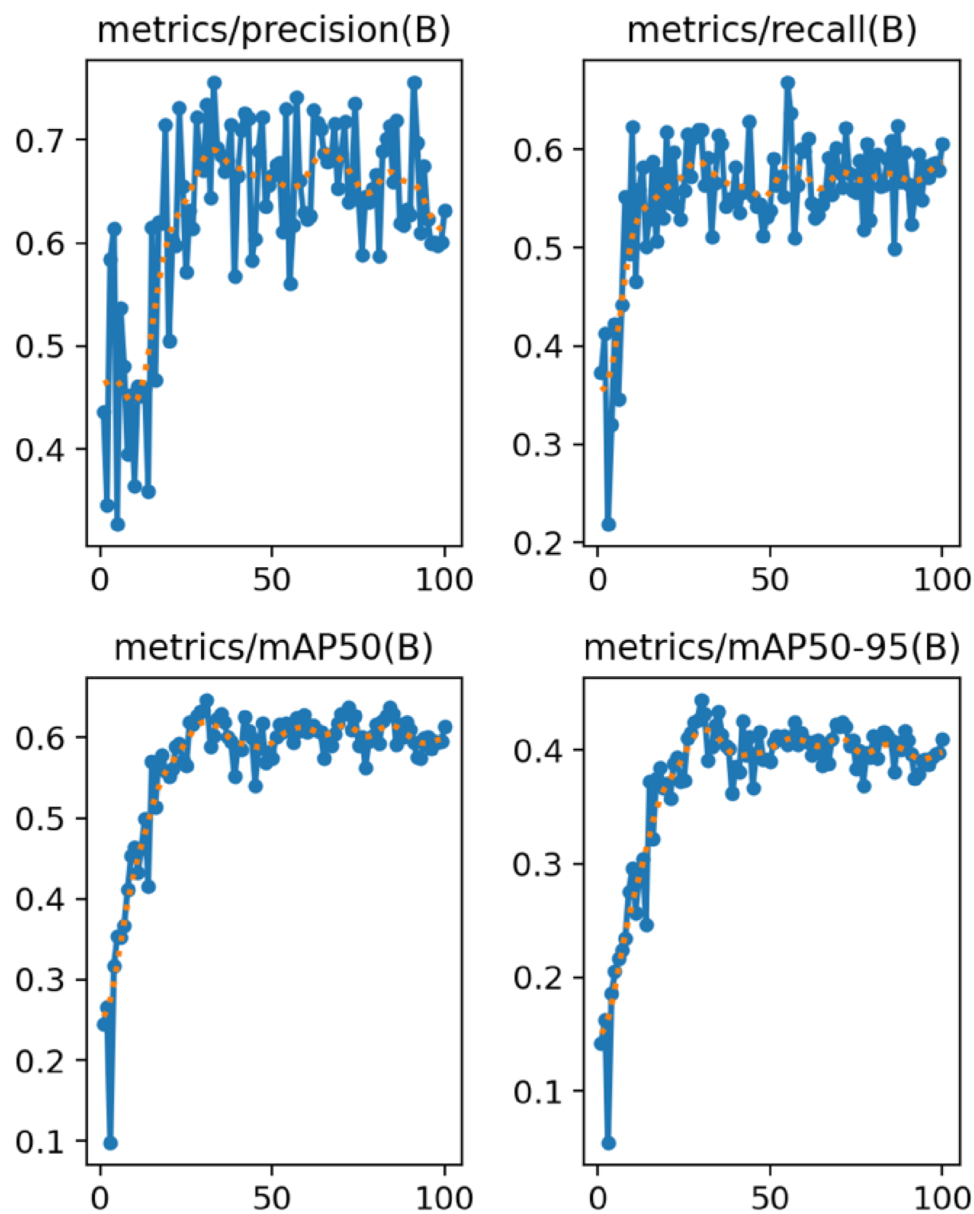

Figure 17: Precision-Recall-mAP

Model	Metrics/precision(B)	Metrics/recall(B)	Metrics/mAP50(B)	metrics/mAP50-95(B
Yolo V8 Small	0.60248	0.54488	0.5646	0.39546
Yolo V8 Medium	0.65002	0.52261	0.54147	0.37731
Yolo V8 Large	0.63182	0.60589	0.61353	0.40979
RT-DETR	0.61839	0.57952	0.59149	0.38697
Faster RCNN	0.491		0.1847 (mAP)	0.3561

Figure 18: Precision-Recall-mAP
Results

8.1 Future work

Several domains of future dental imaging will be in an advantaged situation in particular the notions of development and tracking of superior models as owing to the scopes and diversification of datasets accessible for the task. Expanding the list of dental conditions and image acquisitions, including various, uncommon, and challenging, can improve the models' efficacy and versatility. Further, the extension of such advanced models to novel domains or using dental x-rays to estimate age for chronological age could be useful in forensic dentistry and pediatric dentistry. For increasing the convergence accuracy and mesh refinement, diverse data preprocessing methods like normalization, augmentation and combination of multiple imaging modalities like intraoral scanning with x-rays. These techniques are bound to overcome some of the issues that characterise variability in image quality and variation of anatomic appearances from one patient to another, hence improving the reliability of diagnosis. In conclusion, the following aspects will improve offering proper diagnosis to the patients, thus improving the quality of the care facilities for dental problems.

8.2 Aknowledgement

I would like to thank Mr. Vladimir Milosavljevic for guiding us throughout the research and for all the valuable inputs. I would also like to thank all the authors whose references were used in this research

References

1. Shafi, I., Fatima, A., Afzal, H., Díez, I. de la T., Lipari, V., Breñosa, J., & Ashraf, I. (2023). A Comprehensive Review of Recent Advances in Artificial Intelligence for Dentistry E-Health. *Diagnostics, 13*(13), 2196. [Online] Available at: https://www.mdpi.com/2075-4418/13/13/2196 [Accessed 28th June 2023]. DOI: 10.390/diagnostics13132196.

2. Huang, C., Wang, J., Wang, S., & Zhang, Y. (2023). A review of deep learning in dentistry. *Neurocomputing*. [Online] Available at: https://doi.org/10.1016/j.neucom.2023.126629 [Accessed 2023].

3. Rahulsinh, B. C., Shah, T. V., Shah, D. H., Gohil, T. J., Oza, A. D., Jajal, B., & Saxena, K. K. (2023). An overview of image processing for dental diagnosis. *Innovation and Emerging Technologies, 10.*

4. Qayyum, A., Tahir, A., Butt, M. A., Luke, A., Abbas, H. T., Qadir, J., Arshad, K., Assaleh, K., Imran, M. A., & Abbasi, Q. H. (2023). Dental caries detection using a semi-supervised learning approach. *Scientific Reports, 13*(1), 749. https://doi.org/10.1038/s41598-023-27808-9.

5. Latke, V., & Narawade, V. (2023). A New Approach towards Detection of Periapical Lesions using Artificial Intelligence. *Grenze International Journal of Engineering and Technology, 9*(2), 51. https://doi.org/10.12345/gijet.2023.9.2.51.

6. Karaoglu, A., Ozcan, C., Pekince, A., & Yasa, Y. (2022). Numbering teeth in panoramic images: A novel method based on deep learning and heuristic algorithm. *Journal of Engineering Science and Technology, 17*(3), 101316. https://doi.org/10.1016/j.jestch.2022.101316.

7. Ali, M. A., Fujita, D., & Kobashi, S. (2023). Teeth and prostheses detection in dental panoramic X-rays using CNN-based object detector and a priori knowledge-based algorithm. *Scientific Reports, 13*(1), 16542

8. Abramson, Z., Oh, C., Wells, M., Choudhri, A. F., & Whitehead, M. T. (2024). CT and MR Appearance of Teeth: Analysis of Anatomy and Embryology and Implications for Disease. *Journal of Clinical Medicine, 13*(5), 1187. https://doi.org/10.3390/jcm13051187.

9. You, W., Hao, A., Li, S., Wang, Y., & Xia, B. (2020). Deep learning-based dental plaque detection on primary teeth: a comparison with clinical assessments. *BMC Oral Health, 20*(1), 141. https://doi.org/10.1186/s12903-020-01114-6.

10. Pedro, A. R., Dias, M. B., Laranjo, L., Cunha, A. S., & Cordeiro, J. V. (2023). Artificial intelligence in medicine: A comprehensive survey of medical doctor's perspectives in Portugal. *PLOS ONE*. [Online] Available at: https://doi.org/10.1371/journal.pone.0290613 [Accessed 7 August 2024].

11. Bouchahma, M., Hammouda, S., Kouki, S., Alshemaili, M., & Samara, K. (2019). An automatic dental decay treatment prediction using a deep convolutional neural network on X-ray images. In *2019 IEEE/ACS 16th International Conference On Computer Systems And Applications (AICCSA)* (pp. 1-4).

12. Karamifar, K., Tondari, A., & Saghiri, M. A. (2020). Endodontic Periapical Lesion: An Overview on the Etiology, Diagnosis and Current Treatment Modalities. *European Endodontic Journal*. [Online] Available at: https://doi.org/10.14744/eej.2020.42714 [Accessed 7 August 2024].

13. Sáenz Aguirre, M., Gómez Muga, J. J., Antón Méndez, L., & Fornell Pérez, R. (2022). CT Findings for Dental Disease. *Radiología, 64*, 573–584. Available at: https://doi.org/10.1016/j.rx.2021.06.007 [Accessed 7 August 2024].

14. Ossowska, A., Kusiak, A., & Świetlik, D. (2023). Artificial Intelligence in Dentistry. *Diagnostics, 13*(13), 2196. Available at: https://doi.org/10.3390/diagnostics13132196 [Accessed 7 August 2024].

15. Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machines, 5*(4), 83. Available at: https://doi.org/10.3390/make5040083 [Accessed 7 August 2024].

16. Wang, S., Jiang, H., Li, Z., Yang, J., Ma, X., Chen, J., & Tang, X. (2024). PHSI-RTDETR: A Lightweight Infrared Small Target Detection Algorithm Based on UAV Aerial Photography. *Drones, 8*(6), 240. Available at: https://doi.org/10.3390/drones8060240 [Accessed 7 August 2024].

17. Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*. Available at: https://doi.org/10.1109/JPROC.2023.3238524 [Accessed 7 August 2024].

18. Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimedia Tools and Applications, 82*(10), 9243-9275. Available at: https://doi.org/10.1007/s11042-022-13644-y [Accessed 7 August 2024].

19. Kokomoto, K., Kariya, R., Muranaka, A., Okawa, R., Nakano, K., & Nozaki, K. (2024). Automatic dental age calculation from panoramic radiographs using deep learning: a two-stage approach with object detection and image classification. *BMC Oral Health, 24*, 143. Available at: https://doi.org/10.1186/s12903-024-03928-0 [Accessed 7 August 2024].

20. Abu Tareq, M.I., Islam, M.S., Rafa, N.S., Chowdhury, T., Ahmed, S., Farook, T.H., Mohammed, N. and Dudley, J., 2023. Visual diagnostics of dental caries through deep learning of non-standardised photographs using a hybrid YOLO ensemble and transfer learning model. *International Journal of Environmental Research and Public Health*, 20(7), p.5351. Available at: https://doi.org/10.3390/ijerph20075351 [Accessed 7 August 2024].

Enhancing Early Detection of Dental Caries through Transfer Learning Techniques in Dental Radiography

Himani Sharma

x22224815

Q1: What is a novelty in this work?

The originality of this work is built on the use of novel and transfer learning models: YOLO V8, Faster R-CNN, and RT-DETR, for the challenging task of early identification and characterization of common dental abnormalities including caries, deep caries, impacted teeth, and periapical lesions from panoramic X-rays. Evaluations in differentiation cannot critically rely only on manual analysis of diagnostic results due to discrepancies that may be occasioned by human mistakes, variation in approaches used Unlike this study's implementation of data analytics enhanced by AI tools to improve the precision of the diagnoses that are given. In particular, the research is capable of leveraging transfer learning to transform pre-trained models to perform well within the dental area, which minimizes the requirement of data and computation power compared to the training of models from inception. This makes the process of diagnosis faster while at the same time makes sure that the models are resilient when it comes to clinical setting. However, the central focus of the research as issues significant to dental care and critical for further analysis and intervention remains a high priority yet underdiagnosed in early stages a tool is provided here by this research that would definitely ease the efforts of the health care providers in terms of early identification and more effective intervention resulting in a better patient outcome. The work is most significant in dental radiography as it adopts the modern AI methodologies in the specialized area concentrating on simple but often neglected diseases of teeth, a significant development in AI implementation in healthcare.

The study focuses on categorizing dental problems into four classes: caries, deep caries, impacted teeth and periapical lesion. It illustrates the chances and issue-oriented therapy of caries that is one of the widespread oral diseases; the process of eliminating the softened paint and dentin with the aid of the dental drill and filling the resulting void with metal alloys, glass-polymeric cement, or polymer composite material. The object detection architectures used in this work not only detect these dental issues but also locate them on the images Hence they are not mere image classification as many people may perceive them. Some of the most common image detection architectures include regional CNN, Faster R-CNN as well as you Only Look Once (YOLO) and the latter improves the sharpness of the dental images even under situations when there are food particles or other interferences in the mouth.

Q2: What is the state of the art in this specific task? It's critical for research and could justify your choice of a particular modeling approach.

The landscape of object detection has evolved significantly, with several advanced models representing the cutting edge of performance in various aspects:The landscape of object detection has evolved significantly, with several advanced models representing the cutting edge of performance in various aspects:

1. YOLOv8: YOLOv8 (You Only Look Once version 8) is the latest model which is present in YOLO series Models and it is known for its high speed and accuracy as well. The YOLOv8 family

includes: YOLOv8s (small): This version takes less time for inference and is resource-friendly and makes it ideal for situations where the amount of computational power is limited. It is most beneficial in environments where real-time has to be processed on comparatively low-hardware devices, for example, in an android or even an embedded environment.

YOLOv8m (medium): This model is an intermediate one giving better accuracy compared to YOLOv8s models while inferring with relatively high speed. It is ideal for any situation when speed is an important factor as well as the selectivity of the detector; for example, in video surveillance systems when medium computational power is available.

YOLOv8l (large): YOLOv8l is ideal for use when precision and accurate detection is most critical in an application although this comes at the cost of higher computational required and slower runtime. This model is best suited for applications where maximum precision of image perception is required, for instance in high accuracy industrial robot vision inspection, or in highly sensitive surveillance systems.

2. RT-DETR: RT-DETR is referred to as Real-Time Detection Transformer which is a substantial progress in transformer-based object detection. In doing so, we introduce RT-DETR, which is improved from the DETR model with better convergence speed and faster inference time. It also adapts to the transformer architecture to understand associations and dependencies between different data within images making it very efficient for use in applications where timing is critical, such as; auto-driving or real-time video analysis.

3. Faster R-CNN: Faster R-CNN is still one of the best go-to options for object detection because of the higher accuracy. In the first stage, it employs the Region Proposal Network (RPN) to create potential object areas; at the second stage, it applies Fast R-CNN detector to define and classify them. Faster R-CNN provides slightly less accuracy compared to newer models such as YOLOv8, however, it is more accurate as opposed to other models and therefore useful when providing high levels of accuracy is more important than speed at which the application runs. It is mainly applied in such situation as in diagnosing the images of human bodies or detailed scene recognition because the object detection here must be very accurate.

Q3: You have used different data splits, 80:20 and 85:15. Can you explain why you chose these splits?

Selecting 80:20 and 85:15 data split comes from the need to both train and test the models so as not to overtrain the model. Here's a detailed explanation for each:Here's a detailed explanation for each:

80:20 Data Split

In this approach of data splitting the training data set is composed of 80% while the test data set is composed of 20%. This ratio is used as it offers many amounts of data for training while saving a great amount for testing. This enables the model's evaluation on unseen data, which is an important way of determining the ability of the model to generalize. This split enables the model to have data to learn from especially so if the problem is complex or the data-set very large. It also helps generate a test set having a measure of assessing on which type of data the model is perfect for. This is quite helpful in maintaining a proper trade off between model training and its subsequent testing. For experiments with YOLOv8 Small and Medium and RT-DETR, it is possible to divide the data into such a split to use the test data to check the general performance of the model because such data should be quality checked to be a reasonable sample of the overall data. When there is strong belief that training data is sufficient and accurate then using cross validation can be handy, especially in view of the fact that one wants to use reasonable quantity of data in testing the model.

85:15 Data Split

Here the data is divided where 85% is used to train the model while 15% is used for testing the model. This ratio is usually arrived in a way that can allow more data to be fed to the model especially when working with lower dataset or dataset which needs the model to be fed with many samples. This split increases the number of training data and contributes to the model learning from different examples so as to perform well on new data. This is especially helpful for models such as YOLOv8 (Large) or Faster R-CNN where having a bigger training set improves the model's performance and stability particularly as it applies to the more resource-hungry models which require more training. It is beneficial when the set of data to train the model is not large or, there is the need for fine-tuning and the model needs more instances to learn from in the next training set. This means that the model is able to get a better accuracy of the predictions, as it is demonstrated in the case of YOLOv8 (Large), which received higher results for this kind of data division due to its efficiency in training on a large number of samples.

Q4: With transfer learning, there are other models that can be used to enhance the model deployed. Can you suggest other transfer learning models that could be used?

For further improvement of the existing object detection system different models of transfer learning like EfficientDet, RetinaNet, CentreNet, Cascade R-CNN, and Swin Transformer can be incorporated.

- **EfficientDet:** EfficientDet is an extension of the EfficientNet architecture which applies the compound scaling methods that it uses to scale up all the three components; the depth, the width and the resolution of the architecture in a holistic manner. This makes it possible to achieve a high accuracy in price prediction while at the same time ensure that it does not take too many resources during the computation; therefore it is suitable in scenarios that require high performance but at the same time the available computational resources are limited. It is mainly designed to perform well in real world circumstances while maintaining working expenses at a minimum.
- **RetinaNet:** Another valuable option is RetinaNet which performs especially to handling the class imbalance issue that is common in object detection. Its focal loss function assists to enhance the capacity of the mechanism by lessening the loss incurred to examples that are easily detected and giving more concentration to intricate cases which may comprise petite objects or objects ensconced densely. Therefore, RetinaNet is particularly effective in object detecting in scenes that have many classes of objects and or a high density of objects in the scenes to be analyzed.
- **CenterNet:** CenterNet, they directly estimate object centers and their scales unlike the conventional methods of Estimating Object Proposals using Bounding Box Regression. The above model has been made in a way that it can work perfectly with all manner of sizes and shapes of the objects. That is why by working with the center points and dimensions of the objects it is possible to get rather accurate detections even in the case of big variations in sizes of the objects, which is why CenterNet might be considered rather perspective for different sets of images.
- **Cascade R-CNN:** Cascade R-CNN used a multi-stage structure for the detection that improves its outcomes stage by stage. This makes the iteration more accurate because it concentrates on the harder examples in the next stages; thus, it's more suitable for application in complex scenarios where the objects are hard to detect since they are occluded or in cluttered environments. The cascade mechanism aid in attaining high-quality detections since it enhances the model in detecting complicated instances incremental.
- **Swin Transformer:** Swin Transformer incorporates the hierarchical feature extraction with the windowed based self-attention mechanism, thereby bridging the gap of transformers and efficiency. Since it effectively incorporates hierarchical representations and adaptive computation, it achieves high precision and a high level of performance in several detection benchmarks, suitable for high-precision and high-performance applications.