Tooth Decay Detection Using Different YOLO Algorithms

CSE498R report submitted in partial fulfillment of the requirements for the degree

of

Bachelor of Science in Computer Science and Engineering

by

Md Sazid Ahmed Tonmoy
ID: 1911498042

Under the guidance of

Dr. Sifat Momen

Associate Professor

DEPARTMENT OF ELECTRICAL COMPUTER ENGINEERING
NORTH SOUTH UNIVERSITY
Bashundhara, Dhaka-1229, Bangladesh
Summer 2022

Department of Electrical & Computer Engineering
North South University
Bashundhara, Dhaka-1229, Bangladesh

DECLARATION

It is hereby acknowledged that:

a. No illegitimate procedure has been practiced during the preparation of this document.

b. This document does not contain any previously published material without proper citation.

c. This document represents our own accomplishment while being Undergraduate Students in the North South University.

Sincerely,

Md Sazid Ahmed Tonmoy
ID: 1911498042

Department of Electrical & Computer Engineering
North South University
Bashundhara, Dhaka-1229, Bangladesh

Approval

This is to certify that the CSE498R report entitled Tooth Decay Detection Using Different YOLO Algorithms, submitted by Md Sazid Ahmed Tonmoy (ID: 1911498042) is undergraduate students of the Department of Electrical Computer Engineering, North South University. This report partially fulfil the requirements for the degree of Bachelor of Science in Computer Science and Engineering on September 11, 2022, and has been accepted as satisfactory.

Date:

Dr. Sifat Momen
Associate Professor
Department of Electrical Computer Engineering
North South University
Dhaka, Bangladesh

Dr. Rajesh Palit
Professor and Chair
Department of Electrical Computer Engineering
North South University
Dhaka, Bangladesh

Department of Electrical & Computer Engineering
North South University
Bashundhara, Dhaka-1229, Bangladesh

Acknowledgements

First and foremost, I would want to thank Almighty Allah for providing me the strength to carry out my duties and finish the report.

As a component of the Bachelor of Science (BSc) curriculum, the research program is highly beneficial in bridging the gap between academic knowledge and practical experience. Through the theoretical knowledge, this report has been created to have a practical and professional experience.

I also want to express my sincere appreciation to all the lecturers who helped me by giving me the technical know-how and spiritual support I needed to finish the project thoroughly.

I must express my gratitude to my respected faculty member Dr. Sifat Momen for his undivided attention and assistance in reaching this goal. In addition, I am eternally thankful to the ECE department at North South University for giving me with a course like CSE 498 that allowed me to work on this project in depth and bring it to life as I had imagined.

I appreciate the moral and practical support my family and friends have given me as I've carved out this endeavor.

Abstract

Cavities, commonly known as tooth decay, are portions of the hard surface of teeth that have been irreversibly damaged. It eventually turns into tiny gaps or holes. It can cause infection, pain, and tooth loss if not treated appropriately. For developing and underdeveloped countries, regular dental appointments can be costly. This study aimed to develop and evaluate the performance of a tooth decay detection system that used a deep learning technique based on a convolutional neural network (CNN) to detect tooth decay from oral photographs. We used a collection of 233 pictures of teeth with cavitation. Then, we augmented the dataset and used Roboflow to manage our dataset. The input was clear photos of affected teeth on a white background. After some pre-processing, the dataset was trained on three separate object detection models – YOLOv4tiny, YOLOv5s, and YOLOv6. All of them were evaluated by mean average precision. The mAP@.5 for YOLOv4tiny, YOLOv5s, and YOLOv6 are respectively 98.68%, 98.9%, and 99.25%.

Index Terms: Dataset, tooth decay detection, deep learning, transfer learning.

Table of Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

Tooth decay is one of the most widespread health problems in the world. It occurs when a tooth's enamel is damaged. Cavities in the teeth caused by tooth decay can lead to tooth loss. According to a recent study, tooth decay has surpassed heart disease as the most frequent health problem worldwide, affecting almost 34.1% of the population [2]. People in many parts of the world have limited access to dental specialists. Since caries is not life threatening, many patients with untreated caries wait until it is too late, when major complications have already occurred and treatment is too expensive. Dental cavities that go untreated can lead to pulpitis and periapical disorders [3]. However, tooth decay can be halted and the decay process reversed if diagnosed early enough. The enamel has a self-healing property. As a result, early detection of tooth decay is an important factor of treatments to prevent dental caries. It may make dental treatment more affordable for persons with lower and medium incomes.

Artificial Intelligence (AI) is becoming increasingly popular and widely used in medicine to diagnose and treat patients more rapidly and accurately. Deep learning (DL), an artificial intelligence (AI) method, has been used to automate decision-making processes in numerous clinical dental situations in recent years [4]. By autonomously learning from datasets containing human annotations from dental specialists, the approach, which consists of multi layer ConvNets, has already showed promising accuracy on unforeseen data.

The focus of this study was on identifying tooth decay in three phases. The three phases are visible change without cavitation, visible change with micro-cavitation, and visible change with cavitation. Each phase is distinct in terms of personality, patterns, and shapes. The characteristics of the stages are described as follows and illustrated in Figure 1:

1. Visible change without cavitation: This is the earliest stage of tooth decay, when a lesion forms on the tooth. It causes a slight darkening of the tooth's surface, which is usually white or brown [1].

2. Visible change with micro-cavitation: In this stage, demineralization continues and the tooth enamel (the uppermost layer of the tooth's structure) begins to break down.

3. Visible change with cavitation: In this stage, the dentin layer of the tooth is impacted as the tooth decay proceeds. Bacteria get inside the decaying pulp and causes infection.

Figure 1.1: (a) Visible change without cavitation, (b) Visible change with micro-cavitation, (c) Visible change with cavitation

This paper offers a method for accurately predicting and classifying the three phases of tooth decay using deep learning techniques. For the automatic detection of dental decay from oral photos, we constructed a deep Convolutional Neural Network. The model classifies the presence of dental decay in a given image and uses bounding boxes to locate the findings. We used three different YOLO object detection models to train the dataset. In section V, the comparison study of the three has been analyzed for a better representation and understanding of the trained model's efficiency and accuracy. The rest of the paper is set out as follows: The second section covers relevant works. Section III dives into the model's training phases. Experimental setup is covered in Section IV. Section VI concludes with some ideas for the future.

Chapter 2

LITERATURE REVIEWS

An overview, a summary, and an assessment of the state of knowledge in a particular field of study make up a literature review. It could also highlight methodological concerns and make recommendations for further study.In this chapter, we have reviewed some papers related to our work.

2.1 Deep Learning-Based Dental Plaque Detection on Primary Teeth: A Comparison With Clinical Assessments

Wenzhe You and his team aimed to train a deep convolutional neural network (CNN) to detect caries lesions using DeepLabV3+. 886 intraoral images of primary teeth were utilized to train a conventional neural network (CNN) architecture. 98 intraoral pictures of primary teeth were evaluated by the AI model to verify clinical viability. Additionally, digital camera images of teeth were taken. A skilled pediatric dentist evaluated the images and noted the plaque-containing areas. The plaque-containing sites were then detected after the application of a plaque-disclosing agent. To assess the consistency of the manual diagnosis, the dentist drew the plaque area on the 98 digital camera photographs once again after one week. To assess the diagnostic effectiveness of each method based on lower-resolution pictures, 102 intraoral photographs of primary teeth were annotated to signify the plaque regions acquired by the AI model and the dentist. The degree of detection accuracy was measured using the mean intersection-over-union (MIoU) metric.

To start, they used the visual object classes dataset to pretrain the fundamental DeepLab network and derive the initial weights using transfer learning methods. Second, they used their picture dataset of primary teeth, which includes images of 886 primary teeth before and after employing a dental plaque-disclosing agent, to train a DeepLabV3+ model. To enable the AI model to compare the findings and learn from its errors, the dental plaque detected by the AI model was compared with the actual dental plaque regions.

The MIoU for recognizing plaque on the examined dental pictures was 0.726 ± 0.165. The dentist's MIoU when first diagnosing the 98 photographs obtained by the digital camera was 0.695 ± 0.269 and 0.689 ± 0.253 after one week. The AI model had a higher MIoU (0.736 ± 0.174) than the dentist, and the outcomes were unchanged after one week. The MIoU was 0.652 ± 0.195 for the dentist and 0.724 ± 0.159 for the AI model when they evaluated the 102 intraoral pictures. A paired t-test revealed no statistically significant difference between the human expert and AI model in the ability to identify dental plaque on primary teeth (P >.05). [5]

2.2 Deep Learning Application in Dental Caries Detection Using Intraoral Photos Taken by Smartphones

Mai Thi Giang Thanh and his team worked with Intraoral Photos Taken by Smartphones to train several deep convolutional neural network (CNN) to detect dental caries. The objective of this work was to use a deep learning system to diagnose the phases of smooth surface caries using photos from a smartphone. Materials and procedures A training dataset of 1902 images of teeth's smooth surface acquired using an iPhone 7 by 695 individuals was used. To identify early caries lesions and cavities, four deep learning models—You Only Look Once version 3 (YOLOv3), RetinaNet, Faster Region-Based Convolutional Neural Networks (Faster R-CNNs), and Single-Shot Multi-Box Detector (SSD)—were examined. The International Caries Categorization and Management System (ICCMS) classification of a dentist's diagnosis based on an image inspection served as the reference standard.

The two evaluated models with the highest sensitivity for cavitated caries were YOLOv3 and Faster R-CNN, with 87.4% and 71.44%, respectively. For visibly non-cavitated samples, these two models' sensitivity levels were only 36.9% and

26%, respectively (VNC). For cavitated caries and over 71% for VNC, the specificity of the four models was above 86%. For the clinical diagnosis of dental caries using smartphone photos, the YOLOv3 and Faster R-CNN models showed promise. The new work offers a rudimentary understanding of how AI may be applied in clinical settings once it has been developed in the lab. [6]

2.3 Detection and Diagnosis of Dental Caries Using a Deep Learning-Based Convolutional Neural Network Algorithm

Jae-Hong Lee and his team worked with dental images to evaluate the efficacy of deep CNN algorithms for detection and diagnosis of dental caries on periapical radiographs. A training and validation dataset (n = 2400 [80%]) and a test dataset (n = 600 [20%]) were created from a total of 3000 periapical radiography images. For preprocessing and transfer learning, a GoogLeNet Inception v3 CNN network that has already been trained was employed. For detection and diagnostic performance of the deep CNN algorithm, the diagnostic accuracy, sensitivity, specificity, positive predictive value, negative predictive value, receiver operating characteristic (ROC) curve, and area under the curve (AUC) were calculated. Premolar, molar, and premolar and molar models all had diagnostic accuracies of 89.0 percent (80.4-93.3), 88.0 percent (79.2-93.1), and 82.0 percent (75.5-87.1), respectively. The AUC on premolar, molar, and combined premolar and molar models for the deep CNN method was 0.917 (95 percent CI 0.860-0.975), 0.890 (95 percent CI 0.819-0.961), and 0.845 (95 percent CI 0.790-0.901). The best AUC was produced by the premolar model, which was far better than that for the other models (P 0.001). This study showed how deep CNN architecture might be useful for identifying and diagnosing dental cavities. In periapical radiographs, a deep CNN algorithm significantly improved performance in identifying dental caries.[7]

2.4 Automated Dental Cavity Detection System Using Deep Learning and Explainable AI

An artificial intelligence system that identifies the existence of dental cavities on photos and visually explains each diagnostic was created to manage tooth cavities. The technique used in this work identifies cavities on pictures of several teeth and four tooth surfaces, unlike earlier systems that could only detect cavities on one removed tooth with one exposed tooth surface. 506 de-identified photos from web sources and willing human subjects were gathered for training. A ResNet-27 design was found to be the most effective using curriculum learning, reaching 82.8% accuracy and 1.0 in sensitivity. The system's diagnosis might also be visually explained using Local Interpretable Model Agnostic Explanation. This technology can clearly explain its diagnosis to users, which is an essential ability used by dentists.[8]

2.5 PaXNet: Dental Caries Detection in Panoramic X-ray using Ensemble Transfer Learning and Capsule Classifier

The proposed model benefits from various pretrained deep learning models through transfer learning to extract relevant features from x-rays and uses a capsule network to draw prediction results. On a dataset of 470 Panoramic images used for features extraction, including 240 labeled images for classification, their model achieved an accuracy score of 86.05% on the test set. As long as the difficulties of employing Panoramic x-rays of actual patients are taken into consideration, the resultant score reveals satisfactory detection performance and an increase in caries detection time. Their model achieved recall scores of 69.44% and 90.52% for moderate and severe caries lesions, respectively, in the test set of photos containing caries lesions, demonstrating that it is easier to detect severe caries spots and that effective mild caries identification requires a more robust and bigger dataset. This work is a step toward creating a completely automated effective decision support system to help domain specialists, especially in light of the originality of the current research study's use of panoramic photographs.[9]

Work	Dataset Features	Dataset Size	Algorithm Selection	Evaluation	Results
Deep learning Based Dental Plaque Detection on Primary Teeth: A Comparison with Clinical Assessments	Digital camera images of teeth, Annotated plaque area, Dataset evaluated both human expert and AI	886 training data , 102 testing data	DeepLabV3+	(MIoU) metric.	0.724 ± 0.159
Deep Learning Application in Dental Caries Detection Using Intraoral Photos Taken by Smartphones	Smartphone photos taken with Iphone7, Images of teeth's smooth surface, Annotated area	1902 training dataset	YOLOv3, RetinaNet, Faster R-CNNs, and SSD	Accuracy, Sensitivity	Highest Accuracy YOLOv3 74%
					Highest Sensitivity R-CNN 87.4%
Detection and Diagnosis of Dental Caries Using a Deep Learning Based Convolutional Neural Network Algorithm	Images of dental caries on periapical radiographs, Annotated area, Premolar, molar, and premolar and molar are the classes	3000 images, 2400 training data, 600 testing data	GoogLeNet Inception v3	Accuracy	Premolar 89%
					Molar 88%
					premolar and molar 82%
Automated Dental Cavity Detection System Using Deep Learning and Explainable AI	De-identified images from online sources and consenting human participants	506 De-identified images	ResNet-27	Accuracy	Accuracy 74%
					Sensitivity 87.40%
PaXNet: Dental Caries Detection in Panoramic	470 Panoramic images used for features extraction, including 240 labeled images for classification	470 Panoramic images	PaXNet	Accuracy	86.05%

Chapter 3

PROPOSED METHODOLOGY

The major objective of this work is to develop tooth decay detection model that uses deep learning techniques to aid tooth decay identification. The approach adopted in this work is outlined below.

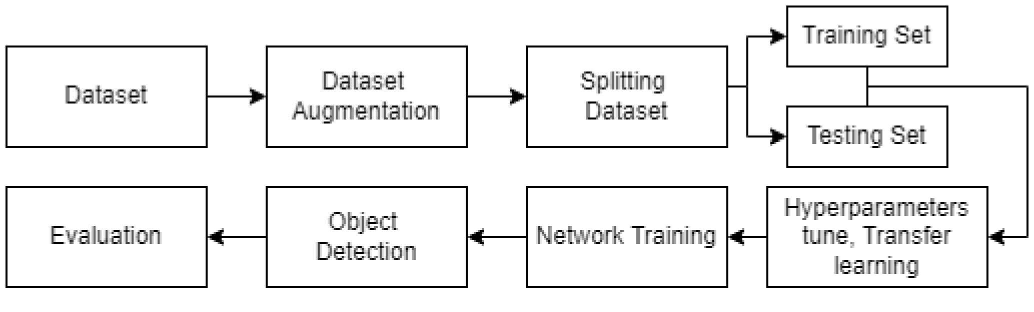

Figure 3.1: Working steps

The process of the entire project is depicted in Figure 2. We started by preprocessing the dataset. The dataset was separated into train and test sets after preprocessing. To develop the object detection model, DL techniques were used on the training set. The test set was then used to assess the models' performance.

3.1 Dataset

3.1.1 Dataset Acquisition

Teeth photos with cavitation, micro-cavitation, and no cavitation were collected from several smartphones to form the tooth decay dataset. There are 233 pictures of teeth in varying stages of decay in the dataset. The dataset was manually annotated by human annotators and encoded in MS COCO format [11].

3.1.2 Dataset Augmentation

We did data augmentation on the existing dataset to increase training and validation accuracy because the dataset only contains 233 photographs. Six augmentation techniques: vertical flip, horizontal flip, 90° rotation, horizontal-vertical flip, average blurring, and raise hue are utilized to turn the dataset into 2618 photos.

3.1.3 Removing Null Values

The dataset was uploaded to Roboflow [10] for image augmentation before training on YOLOv5s. Roboflow found 475 images with Null values after augmentation. Those were taken out of the dataset.

3.1.4 Dataset Splitting

The dataset for YOLOv5s was divided into three parts: a training set with 86% data (2266 images), a validation set with 9% data (231 images), and a test set with 5% data (images).

3.2 Model Training

The dataset was trained in three YOLO object detection models.You Only Look Once is known by the acronym YOLO. This program identifies and finds different things in an image (in real-time). The class probabilities of the discovered photos are provided by the object identification process in YOLO, which is carried out as a regression problem. Convolutional neural networks (CNN) are used by the YOLO method to recognize items instantly. The approach just needs one forward propagation through a neural network to identify objects, as the name would imply. This indicates that a single algorithm run is used to do prediction throughout the full picture. Multiple class probabilities and bounding boxes are concurrently predicted using the CNN. YOLO algorithm works using the following three techniques:

1. **Residual blocks:** The image is initially split into a number of grids. Each grid has a SxS dimension. There are several equal-sized grid cells. Every grid cell will be able to detect items that enter it. For instance, a grid cell will be in charge of detecting an item if its center appears within that cell.

2. **Bounding box regression:** A bounding box is a highlighted outline of an item in a picture. The properties of each bounding box in the image are: Width, Height, Class, and Bounding box center. In order to determine the height, breadth, center, and class of an item, YOLO use a single bounding box regression.

3. **Intersection over union (IOU):** Box overlapping is described by the object detection phenomena known as intersection over union (IOU). IOU is used by YOLO to create an output box that properly encircles the items. The predicted bounding boxes and their confidence scores are the responsibility of each grid cell. If the projected bounding box and the actual box match, the IOU is equal to 1. Bounding boxes that are not equivalent to the actual box are eliminated by this approach.

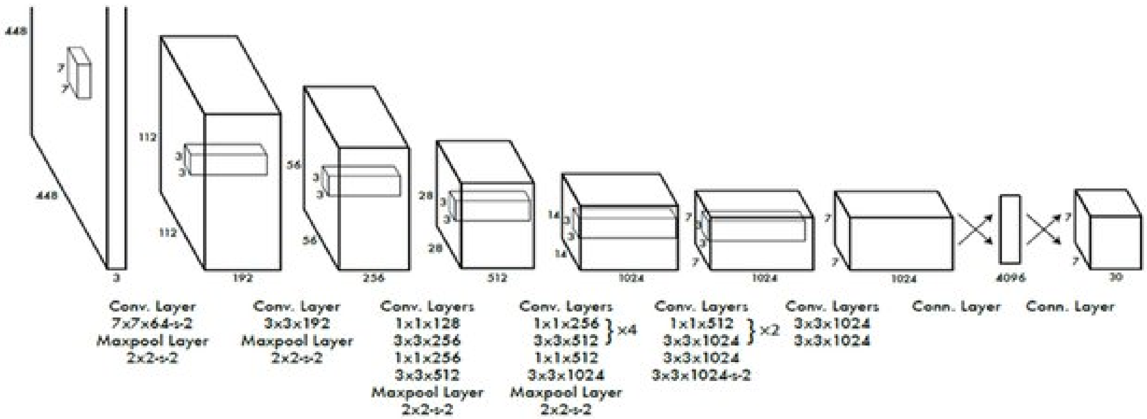

Figure 3.2: Yolo Architecture

Chosen models are: YOLOv4, YOLOv5, and YOLOv6. Model descriptions and how the dataset is trained on the model are described as follows:

3.2.1 YOLOv4

The YOLOv4 algorithm is an enhanced version of the YOLOv3 method. It connects the YOLOv3 head with a CSP darknet-53 [12] classifier and spatial pyramid pooling. It benefits from excellent detection accuracy, precise bounding box positioning, and fast computations. YOLOv4 takes an image as input and compresses features using a backbone of convolutional neural networks. These backbones represent the network's endpoint in picture categorization, and they can be used to make predictions. YOLOv4 is pretrained on ImageNet classification [13]. We used the pretrained YOLOv4 weights and used it to train the model on our custom dataset via transfer learning [11].

Activation Function: Non-monotonic, smooth activation function Mish was chosen due to its minimal cost and qualities such as unbounded above and below, which boost its performance when compared to other often used functions. The Mish function has the following definition:

$$f(x) = x.tanh\left(s\left(x\right)\right)$$

$$s(x) = ln\left(1 + e^{x}\right)$$

S is a softmax activation function.

3.2.2 YOLOv5:

YOLOv5 is available in four different models. Except for the model layer architecture and a few parameters, there are no significant differences between the versions [14]. We chose YOLOv5s and YOLOv5m as we had a limited dataset and trained the models from scratch to detect tooth decay. These two models were trained in two different settings to see which one performed better on our custom object detection. Stochastic Gradient Descent (SGD) optimization was used and the learning rate was 0.01.

Activation Function: YOLOv5 uses Leaky RelU activation function for hidden layers. The use of Leaky ReLU reduces the computation necessary to drive the neural network from growing exponentially. It has the following definition:

$$f\left(y\right) = \left(\alpha y\right); if\left(y < 0\right)$$

$$f\left(y\right) = \left(y\right); if\left(y > 0\right))$$

Sigmoid activation function that was used in the final detection layer. It transforms the input into a value between 0 and 1. The function is defined as follows:

$$Sigmoids(x) = 1/(1 + e^{-x})$$

Loss Function: Binary cross entropy loss function was used. Each of the projected probabilities is compared to the actual class output, which can be either

0 or 1. The score is then calculated, penalizing the probabilities based on their deviation from the predicted value.

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}(y_i * y_{pred} + (1 - y_i) * log(1 - y_{pred})))$$

Where ypred is the ith scalar value in the model output and yi is the corresponding target value.

3.2.3 YOLOv6

A team at the Chinese e-commerce platform business Meituan developed the object identification model known as YOLOv6. The term YOLOv6 is being used by the authors to save space even though the official name is MT-YOLOv6. They have essentially created the model on top of the YOLO (You Look Only Once) architecture and assert various advantages and new techniques over previous YOLO family models. PyTorch was used to create this framework. It adopted the decoupled head structure, taking into account the balance between the representation ability of the operators and the computing overhead on the hardware.It also adopted three strategies:

1. **Anchor-free paradigm:** The anchor-free paradigm has gained popularity in recent years because of its excellent generalizability and straightforward code logic. The team discovered that the Anchor-free detector has a 51%

increase in speed when compared to previous techniques.

2. **SimOTA Tag Assignment Policy:** The researchers employed the SimOTA method, which dynamically distributes positive samples to increase detection accuracy, to acquire high-quality positive samples.

3. **SIoU bounding box regression loss:** To oversee the network's learning process, YOLOv6 uses the SIoU bounding box regression loss function. Through the addition of a vector angle between necessary regression, the SIoU loss function redefines the distance loss. As a consequence, the detection accuracy is enhanced as well as the regression accuracy.

3.3 Evaluation

We should be prepared with several assessment metrics to examine the classification algorithm in the event of a classification problem. As follows:

1. **Confusion Matrix:** The classification model's accuracy in classifying instances into distinct groups is summarized in a table called the confusion matrix. The model's anticipated label is on one axis of the confusion matrix, while the actual label is on the other. When comparing several models, we may use the confusion matrix to assess how well each one predicted true positives (TP) and true negatives (TN). We chose a model as our basic model if it accurately predicted TP and TN compared to other models.

	Positive	negative
Positive	TP	FP
Negative	FN	TN

Figure 3.3: Confusion Matrix

TP = True Positive (The total number of images that are correctly detected to be positive)

FP = False Positive (The total number of images that are predicted to be positive but actually are negative)

TN = True Negative (The number of images that are accurately predicted to be negative)

FN = False Negative (The number of images that are incorrectly predicted to be negative)

2. **Precision and Recall:** Precision and recall are two metrics used to evaluate classification and retrieval systems' performance. Precision is the percentage of relevant occurrences among all retrieved examples. Recall, also known as sensitivity, is the percentage of recovered instances among all appropriate models. In a perfect classifier, precision and recall are both one.

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

3. **Accuracy:** It is calculated by dividing the total number of correctly categorized instances by the overall number of classified examples. When the importance of each class's prediction error is equal, this measure is crucial. Here, false positives should be addressed more than false negatives.

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

4. **F1 score:** A weighted average of recall and accuracy is the F1 score. False positive and false negative results can occur in accuracy and recall, as is well known, thus both are taken into account. In most cases, the F1 score is more helpful than accuracy, particularly if your class is distributed unevenly. When false positives and false negatives cost about the same, accuracy performs best. It is preferable to include both Precision and Recall if the costs of false positives and false negatives are significantly different.

$$F1Score = \frac{2 * Recall * Precision}{Recall + Precision}$$

5. **Learning Curve:** For algorithms that learn (optimize their internal parameters) gradually over time, like deep learning neural networks, learning curves are frequently employed in machine learning. If maximization is the metric used to measure learning, then higher scores (bigger numbers) signify more learning. Accuracy in categorisation might serve as an example. It is more typical to employ a score that minimizes, like loss or error, where better scores (lower numbers) imply greater learning and a value of 0.0 indicates that the training dataset was learnt properly with no errors. Additionally, a hold-out validation dataset that is separate from the training dataset can be used to test it. An assessment of the validation dataset provides insight into the model's "generalizability."

6. **Learning Curve:** Mean Average Precision (mAP) is a popular metric for measuring how well object identification and segmentation systems perform. A statistic called Mean Average Precision (mAP) is used to assess object detection algorithms like Fast R-CNN, YOLO, Mask R-CNN, etc. Recall values between 0 and 1 are used to determine the average precision (AP) values. We need IOU, Precision, Recall, Precision Recall Curve, and AP in

order to compute mAP [15][16]. The mAP was computed using the confidence threshold. average precision AP

$$= \sum_{k=0}^{k=n-1} [\text{Recall}(k) - \text{Recall}(k+1)$$

$$* \ \text{Precision } (k)]$$

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP(k)$$

Where $\mathrm{AP_K}$ = the AP of class k

n = number of classes

Chapter 4

EXPERIMENTAL SETUPS

After classifying the dataset into three classes: 0 (visible change without cavitation), 1(visible change with microcavitation), and 2(visible change with cavitation), the dataset was fed into the model as shown in Fig. 2. We employed three YOLO object detection models. The experimental setup for YOLOv4, YOLOv5, and YOLOv6 is described below-

4.1 YOLOv4 Setup

Images are fed via convolutional down sampling, then supplied through a succession of layers of dense connection blocks that execute various operations and calculations. The outputs of these blocks were then routed via a spatial pyramid pooling layer to widen receptive fields, and then through an object identification layer to identify the various classes in a picture.

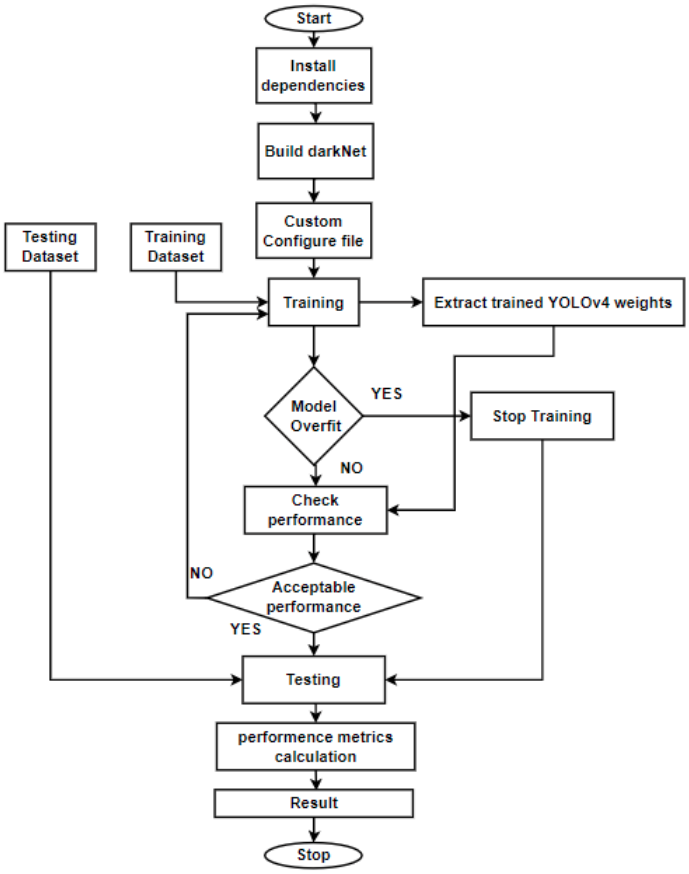

Figure 4.1: Working Flow

Above figure shows the detailed flowchart of design and implementation from splitting dataset to evaluating result. All the necessary dependencies such as YOLOv4, CUDA, NumPy, and Python were installed from respective repositories. DarkNet [12] is a CUDA-based open-source neural network framework designed to support graphics processing units (GPU). A custom configuration file was created from the cloned YOLOv4 repository to build a custom object detector for tooth decay detection. All hyper parameters designed in the development of the custom object detector were detailed in the custom configuration file. The training module was then incorporated with a specific configuration file. Then, the model became ready to be trained with a custom dataset.

Design Constraints and Parameters-Image Dataset

1. Number of Classes: 3
2. Class name: Visible change without cavitation (0), Visible change with micro cavitation (1), Visible change with cavitation (2).
3. Filter Size: 416*416
4. Batch Size: 64
5. Subdivision: 32
6. Number of filters: (3+5)*3 = 24

Hyper parameters design

1. Image Size: 416*416
2. Image channels: 3
3. Kernel Size: 3*3
4. Activation Function: Mish
5. Batch Size: 64
6. Max batches: 3*2000 = 6000
7. Learning rate: 0.001

The trainer's supervision was necessary for each epoch's parameter values, such as mAP, Accuracy, and Precision. To avoid model overfitting, training should be stopped after the given parameter values become constant or have very few changes. After the model was trained, best YOLOv4 weights are extracted which acts as a reference while testing the model on custom testing dataset.

4.2 YOLOv5 and Yolov6 Setup

A similar working flow as seen in fig 4.1 was used to train the dataset in YOLOv5 and YOLOv6 but without darknet. A dataset of 2618 photos was used to train YOLOv5 and Yolov6, with 86% used for training, 9% for validation, and 5% for testing. Both the model development started by installing necessary dependencies such as Python, Pytroch, YOLOv5, CUDA and roboflow from respective repositories. Datasets were uploaded and then exported into the YOLOv5 PyTorch format, which generated api keys for each dataset. It's worth mentioning that the Ultralytics solution requires a YAML file that defines where your training and test data should be stored. The Roboflow export also creates this format for us. 0 indicates visible changes without cavitation, 1 indicates visible changes with micro-cavitation, and 2 indicates visible changes with cavitation in the data.yaml file. Training configuration for YOLOv5 and YOLOv6:

Table 4.1: YOLOv5 and YOLOv6 MODEL CONFIGURATIONS

Required Argument	YOLOv5s	YOLOv5m
Image Size	416x416	416x416
Batch Size	16	32
Epoch	100	100

Training losses and performance data was saved to a log file created before with the —name flag when the model was trained. The weight values were saved in .pt files. On test photos, the best weights was applied and got somewhat improved accuracy in YOLOv5 and YOLOv6.

Chapter 5

EXPERIMENTAL RESULTS

Precision and recall are two metrics used to evaluate classification and retrieval systems' performance. Precision is the percentage of relevant occurrences among all retrieved examples. Recall, also known as sensitivity, is the percentage of recovered instances among all appropriate models. In a perfect classifier, precision and recall are both one. Object detection models like YOLO use the assessment measure mAP (mean Average Precision). To calculate mAP, we needed IOU, Precision, Recall, Precision Recall Curve, and AP.

The YOLOv4 model training took more time compared to YOLOv5 and YOLOv56. The model achieved the overall mAP of 98.68%. Below table shows the overall result on YOLOv4 training.

Table 5.1: YOLOv4 Results

Class	Average Precision	mAP@0.50
Visible change without cavitation	96.61%	98.68 %
Visible change with microcavitation	99.85%	
Visible change with cavitation	100.00%	

For conf thresh = 0.25, precision = 0.97, recall = 0.96, F1-score = 0.97. Again, For conf thresh = 0.25, TP = 285, FP = 8, FN = 12, average IoU = 79.60%. IoU threshold = 50%, used Area-Under-Curve for each unique Recall. Mean average precision (mAP@0.50) = 0.988213, or 98.82%. Total Detection Time: 1 Seconds The YOLOv5s model was trained with 100 epoch and outperformed the YOLOv4 models. The model achieved the precision of 97.8% on overall class. mAP@.5 was 98.9% and mAP@.5:.95 was 76.7%. The classwise result for YOLOv5 is given below:

Table 5.2: YOLOv5 Results

Class	Precision	Recall	mAP@.5	mAP@.5:.95
All	97.8%	98%	98.9%	76.7%
Visible change without cavitation	99.3%	95.8%	98%	72.4%
Visible change with microcavitation	95.3%	98.2%	99.1%	76.1%
Visible change with cavitation	98.9%	100%	99.6%	81.6%

The model YOLOv6 on the other hand, was trained with 100 epoch and outperformed slightly compared to YOLOv5. The model achieved an accuracy of

99.24%. The classwise result for YOLOv6 is given below:

Table 5.3: YOLOv6 Results

Class	Precision @.5:.95	Recall@.5:.95	mAP@.5
ALL	88.4%	91.3%	99.24%
Visible change without cavitation	100%	100%	
Visible change with microcavitation	83.7%	84.8%	
Visible change with cavitation	88.8%	91.8%	

YOLOv6's result is generated as a json file. So, Only graphical results of YOLOv5 is shown below:

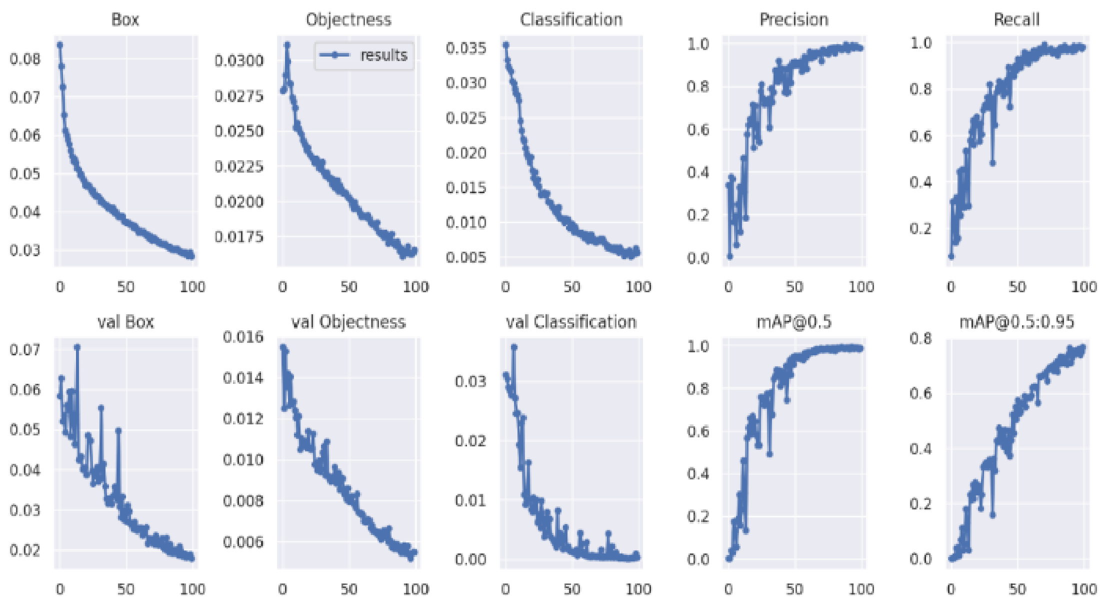

Figure 5.1: Result of training in YOLOv5

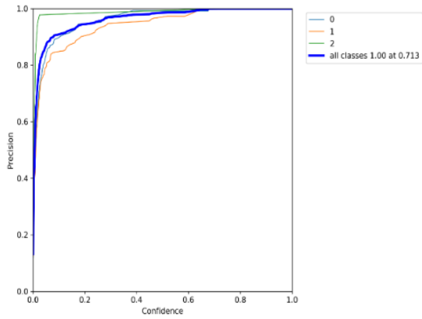

Figure 5.2: Precision curve for YOLOv5

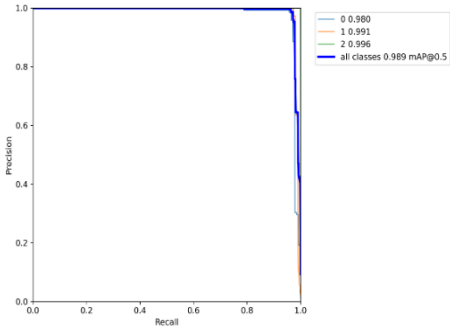

Figure 5.3: Precision-Recall curve for YOLOv5

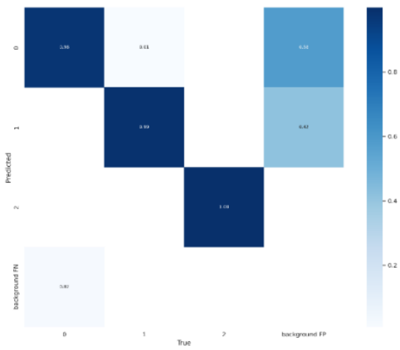

Figure 5.4: Confusion matrix for YOLOv5

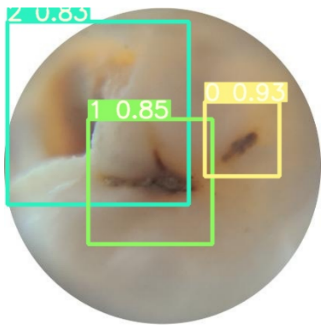

Figure 5.5: Sample Output of YOLOv5 model

All of the above model train in same environment (Google Colab) with almost same condition. So, we can compare their mAP and time in a graph for better visualization.

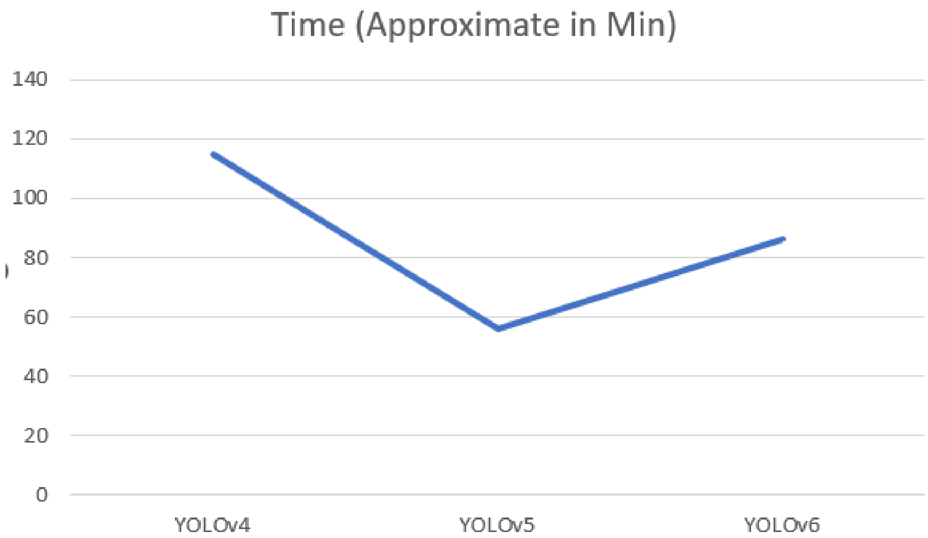

Figure 5.6: Model VS Time

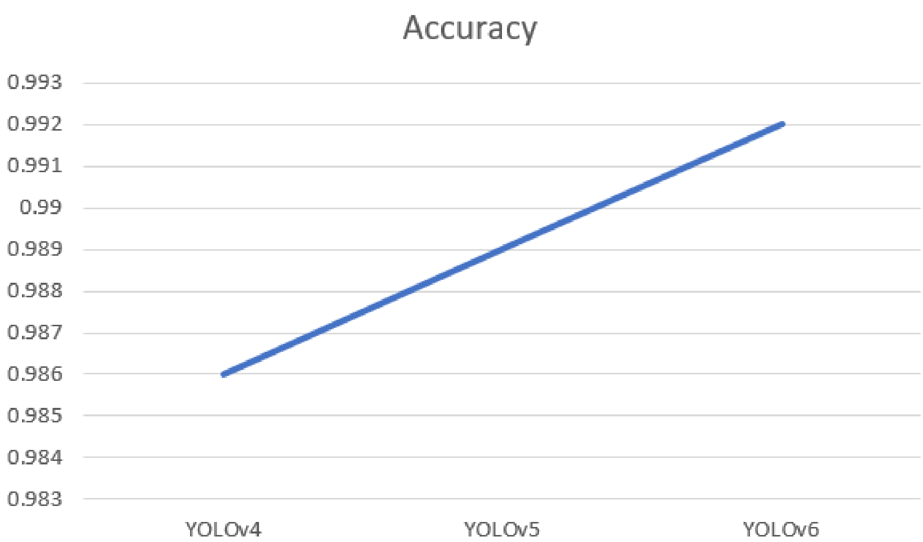

Figure 5.7: Model VS mAP@0.5

Chapter 6

CONCLUSION

Early identification of tooth decay can save costs on dental treatments and even reverse the decay process. A deep learning strategy to detect tooth decay is presented in this paper. The dataset was trained in two different object detection algorithms to see which one performed better. YOLOv6 was more accurate but YOLOv5 was faster in this condition. YOLOv4 was good but YOLOv5 and YOLOv6 was better in every angle. The YOLOv6 model had the highest accuracy of 99.2%, while the YOLOv4 model had the lowest accuracy of 98.6%.

In comparison to the current, CNN-based dental caries classification models, This suggested technique can produce results that are more spectacular while using less computing power. The nicest thing about our suggested approach is how easy it may be applied to various kinds of illness categorization based on medical images. Such a method will change the area of visual illness diagnostics and be of enormous use to medical Experts. Additionally, this work could benefit from the use of ocular image segmentation. Additionally, a system like this would revolutionize the field of diagnosing dental caries and be very helpful to medical professionals. my opinion is that it can still be a valuable model, and there will likely be possibilities to improve it with further research and study in the near future.

Bibliography

[1] B. J. N. D.D.S., "The Stages of Tooth Decay: Progression of a Cavity," 16 6 2017. [Online]. Available: https://www.nordhusdentistry.com/blog/2017/06/16/the-stages-of-tooth-decay-183529. [Accessed 12 5 2022].

[2] Peres, M. A., Macpherson, L. M. D., Weyant, R. J., Daly, B., Venturelli, R., Mathur, M. R., Listl, S., Celeste, R. K., Guarnizo-Herreño, C. C., Kearns, C., Benzian, H., Allison, P., Watt, R. G. (2019). Oral diseases: A global public health challenge. The Lancet, 394, 249–260. https://doi.org/10.1016/S0140-6736(19)31146-8

[3] Ali, A. H., Koller, G., Foschi, F., Andiappan, M., Bruce, K. D., Banerjee, A., Mannocci, F. (2018). Self-limiting versus conventional caries removal: A randomized clinical trial. Journal of Dental Research, 97(11), 1207–1213. https://doi.org/10.1177/0022034518769255

[4] Park, W., Park, J. (2018). History and application of artificial neural networks in dentistry. European Journal of Dentistry,12(4),594-601, https://doi.org/10.4103/

[5] W. You, A. Hao, S. Li, Y. Wang, and B. Xia, "Deep learning-based dental plaque detection on primary teeth: a comparison with clinical assessments," BMC Oral Health, vol. 20, no. 1, May 2020, doi: 10.1186/s12903-020-01114-6.

[6] M. T. G. Thanh, N. Van Toan, V. T. N. Ngoc, N. T. Tra, C. N. Giap, and D. M. Nguyen, "Deep Learning Application in Dental Caries Detection Using Intraoral Photos Taken by Smartphones," Applied Sciences, vol. 12, no. 11, p. 5504, May 2022, doi: 10.3390/app12115504.

[7] J.-H. Lee, D.-H. Kim, S.-N. Jeong, and S.-H. Choi, "Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm," Journal of Dentistry, vol. 77, pp. 106–111, Oct. 2018, doi: 10.1016/j.jdent.2018.07.015.

[8] Bhattacharjee N. Automated Dental Cavity Detection System Using Deep Learning and Explainable AI. AMIA Annu Symp Proc. 2022 May 23;2022:140-148. PMID: 35854730; PMCID: PMC9285146.

[9] Haghanifar, Arman, Mahdiyar Molahasani Majdabadi, and Seok-Bum Ko. ”Paxnet: Dental caries detection in panoramic x-ray using ensemble transfer learning and capsule classifier.” arXiv preprint arXiv:2012.13666 (2020)

[10] Tsung-Yi Lin Michael Maire Serge Belongie Lubomir Bourdev Ross Girshick James Hays Pietro Perona Deva Ramanan C. Lawrence Zitnick Piotr Dollar, ” Microsoft COCO: Common Objects in Context”, pp. 4-6, 2014, https://doi.org/10.48550/arXiv.1405.0312

[11] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, ” YOLOv4: Optimal Speed and Accuracy of Object Detection”, pp. 6-9, 2020, https://doi.org/10.48550/arXiv.2004.10934

[12] ”YOLOv4- An explanation of how it works”, Roboflow Blog,2022. [Online]. Available: https://blog.roboflow.com/a-thorough-breakdown-of-yolov4/.

[13] ”Custom Object Detection with transfer learning with pre-trained YOLO-V4 model”, sandipanweb, 2022. [Online].Available:https://sandipanweb.wordpress.com/2022/01/17/custom-object-detection-with-transfer-learning-with-pre-trained-yolo-v4-model/. [Accessed: 14- May- 2022].

[14] ”YOLOv5 is Here”, Roboflow Blog, 2022. [Online]. Available:https://blog.roboflow.com/yolov5-is-here/

[15] ”IoU a better detection evaluation metric”, Medium, 2022.[Online].Available:https://towardsdatascience.com/iou-a-better-detection-evaluation-metric.

[16] ”Mean Average Precision (mAP) Explained | Paperspace Blog”, Paperspace Blog, 2022. [Online]. Available:https://blog.paperspace.com/mean-average-precision/