Computer vision with universal smartphone microphotography in the detection of tooth tissue destruction

Muntasir Md. Nafis , Sadiqun Nur Ayan , Faisal Osman
*

***Abstract*—Tooth tissue destruction is one of the most common chronic conditions in the world. Until recently, practically everyone had at least one encounter with this pathology over their lifetime. Early identification of dental caries can result in a significant reduction in the rate of dental disease. Clinical applications now have a greater impact on patient care as medical imaging becomes more accessible. Machine learning algorithms for classification and analysis of picture data have recently piqued people's curiosity. We present a new method for detecting and identifying dental decay utilizing images as the dataset and a deep neural network as the strategy in this research. The use of a deep neural network to diagnose dental decay is innovative. This approach was tested on a real dataset and has demonstrated a good performance of detection.**

***Index Terms*—Deep Neural Networks, YOLOv5, Object Detection, mAP.**

I. Introduction

Dental caries is the most common chronic disease on the world. Early lesions remain prevalent in the majority of the population, notwithstanding significant decreases in the prevalence of big cavitated lesions. Conventional caries detection methods that rely on visual inspection and the use of a dental probe are effective for large, clearly visible caries and for those partially obscured but accessible with a handheld mirror. However, it is well recognized that the introduction of new technologies would aid in the diagnosis of dental lesions, which is an essential driver of treatment outcomes.

Cavities are very common. More than nine out of ten people would have had tooth decay at some point. This infection has affected more than a third of 6 year-old children and more than half of 12 year-old children in France. In Canada, 57 percent of children aged 6 to 12 have at least one tooth decay.

In dentistry, clinical photography has a wide range of applications. Their responsibilities include diagnostic and medicolegal goals, patient and professional communication, and education. With the use of smartphones in intra- and extraoral feature evaluation, tele- dentistry has been greatly simplified.

The term "Dental Microphotography" refers to the use of a dental microscope and a digital camera to capture magnified images of a specific intra- or extraoral region of interest. The use of smartphone microphotography in studying carious mineralization changes and histopathological evaluation of oral diseases has been well documented. Google and Apple smartphones, as well as Google and affiliates' Artificial Intelligence (AI) platforms, have recently introduced a variety of AI-driven microphotography applications in medical diagnostics. While digital cameras capture higher-quality images by definition, the lighter weight, increased portability, and constantly improving camera stabilization system have led medical professionals to prefer smartphone-based solutions to medical diagnostic challenges.

Machine learning is defined as the ability to teach an agent how to make decisions based on observations [4]. The action of this agent in the biomedical context is reflected by additional information to assist the dentist in making his decision. In dentistry, computer vision is defined as the application of machine learning to visual data via neural networks in order to recognize images and patterns. This is accomplished by categorizing entire images, detecting objects and ROIs within larger images, and image segmentation.

Tooth decay from dental caries is considered a non-fatal, silent 'pandemic' in that it affects adults and can lead to debilitation if left untreated. The International Caries Detection and Assessment System (ICDAS) has been widely accepted as a reliable standard for diagnosing caries on a global scale. Pitts (2013) In machine learning attempts to automate the predictive diagnosis of dental caries and tooth decay, the ICDAS classification system has been successfully implemented.

II. Relevant Work

Detecting dental caries has traditionally been a visual process based on visual-tactile examination and radiographic examination [1]. Several techniques for detecting dental caries have recently been developed in the literature. Kositbowornchai et al. [2] used images from a charged coupled device (CCD) camera and intra-oral digital radiography to train a neural network to detect artificial dental caries. The main disadvantage of this method is that the system was evaluated using artificial carries, which are completely different from naturally affected ones. Saravanan et al. [3] used histogram and power spectral analysis to develop a new method for detecting dental caries in its early stages. The detection of tooth cavities in this method is based on the region of concentration of pixels in relation to the histogram and the magnitude values in relation to the spectrum. The main disadvantage of this study is that it is solely dependent on the intensity of pixels. Berdouses et al. [4] created a computer-aided automated methodology for detecting and classifying occlusal caries in color photographs. The segmentation of photographic color images is the foundation of this method.

Even though there are many methods for early caries detection, it is still necessary to develop accurate caries detection methods to assist dentists.

The problem with the traditional approach to form recognition is that it is extremely difficult to create a good characteris-

tics extractor that must be readjusted for each new application. Deep learning is a class of methods whose principles have been known since the late 1980s, but whose application has only recently become widespread.

One of the perspectives of deep learning techniques is the replacement of laborious work by algorithmic models of supervised learning, nonsupervised (i.e., not requiring specific knowledge of the problem studied), or techniques of hierarchical characteristic extraction.

The concept is straightforward: the training system consists of a series of modules, each of which represents a processing step. Each module can be trained with adjustable parameters similar to the linear classifiers' weight. The system is trained from start to finish: for each example, all the parameters of all the modules are adjusted to approximate the system's output of the desired output. The term "deep qualifying" refers to the arrangement of these modules in successive layers [5].

To train the system in this manner, it must be known which direction and how much to adjust each module's parameters. This necessitates the computation of a gradient. This gradient is calculated using the back-propagation method, which has been used since the mid-1980s. A deep architecture can be thought of as a multilayer network of simple elements interconnected by training weight, similar to linear classifiers. This is referred to as a multi-layer neural network.

Deep architectures have the advantage of being able to learn to represent the world in a hierarchical manner. There is no need to hand-build a characteristics extractor because all layers can be trained. It will be accomplished through training [6]. Furthermore, the first layers extract some simple characteristics, and the subsequent layers combine to form more and more complex concepts.

III. Methodology

In this section, we will describe the use of deep neural networks with multiple hidden layers to train and detect decays from tooth images. Multiple hidden layers neural networks can be extremely useful in solving object detection problems involving complex data, such as images. Each layer can learn features at a different level of abstraction. To train our model, we used the YOLOv5 algorithm. The sigmoid activation function is used in the final detection layer, while the leaky ReLU activation function is used in the middle/hidden layers.

A. Participant Data

Our dataset consists of 233 microphotography images. As described in the target condition, these images were classified into three machine-learning classifications of tooth decay obtained by photomicrography and labelled in yolo format.

B. Data Transformation

Data augmentation is a common performance-enhancing strategy in which the size and diversity of the data set are increased by applying various transformations to the original images. In our training pipeline, we used the following transformations to augment the data: blurring, contrast limited adaptive histogram equalization, gray coloring, changing brightness, changing contrast, and decreasing image compression.

C. Dataset Patitioning

Bounding boxes in yolo format were used to label the images. The model was given both labelled and unlabelled images as training data. The dataset was divided into a 9:1 ratio (90 percent -10 percent) for training and validation purposes. 90 percent images of the dataset were used as training data to train the model. Rest of the images (10 percent) were used as validation data to evaluate the model's performance on unseen data.

D. Classification Points

Images in our dataset were divided into 3 classes. The classification points are : 1) Visible change without cavitation, 2) Visible change with microcavitation, 3) Visible change with cavitation.

E. Activation Function

In general, activation functions operate on preactivation vectors element by element. In our model, Leaky ReLU activation function is used in middle/hidden layers and the sigmoid activation function is used in the final detection layer.

When the input is less than zero, the Leaky ReLU modifies the function to allow small negative values. When the unit is saturated and not active, the leaky rectifier allows for a small, non-zero gradient. Leaky ReLU function is defined as :

$$f(x) = 1(x < 0)(x) + 1(x >= 0)(x) \tag{1}$$

On the other hand, The Sigmoid Function is also an activation function that is used to add non-linearity to a machine learning model by deciding which values to pass as output and which not to pass. The sigmoid function is defined as :

$$S(x) = \frac{1}{1 + e^{-}x} \tag{2}$$

This function takes an input and outputs values in the range of 0 to 1.

IV. Experimental Setup

300 microphotography photos were captured using a smartphone (Samsung Galaxy S20 5G; Samsung Electronics) equipped with a microscope. As described in the target condition, these photos were classified into four machine-learning classifications. Following full in-person agreement, the three dentists selected and labeled virtual areas of interest (v-ROI). For computer vision object detection, virtual labeling was used within the LabelImg system(Yolov4).

A teaching hospital in Malaysia submitted 300 microphotographic photos of excised permanent human teeth with hidden treatment histories and patient demographics. The sample size (103 to 585 pictures) was within the range of prior reports of similar machine learning applications. The three dentists

examined all of the resulting digital photos for acceptability, and those that did not achieve a = 1.00 interrater agreement were eliminated, leaving 233 images that were entirely agreed upon.

V. Experimental Results and Discussion

In this section, we present the findings of testing experiments we conducted to evaluate our deep neural network's performance to detect decays in microphotographic dental images. Here are some examples of the images from the dataset.

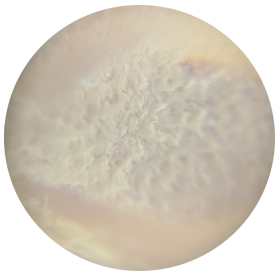

Fig. 1. Visible change without cavitation

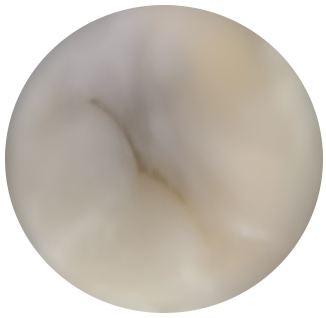

Fig. 2. Visible change with microcavitation

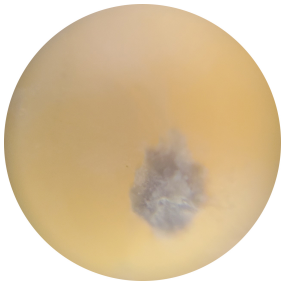

Fig. 3. Visible change with cavitation

A. Without Augmentation

At first, the model was trained using YOLOv5 on 233 images from the tooth decay dataset. The aim of our work was to detect decays in photomicrographic dental images. 90 percent of the dataset was used as training data to train the model and the remaining images were used as validation set to evaluate the model's performance on unseen data. There were three possible labels for each image, visible change with cavitation, visible change without cavitation, visible change with microcavitation.

These are three images from three different classes.

For this experiment, 400 epochs were performed to train the model using the training data. For the 369^{th} epoch, the best result had been achieved. For the best case scenario, 64% Mean Average Precision, 71% Precision and 56% recall value were obtained.

A confusion matrix was generated to assess the overall performance of the model.

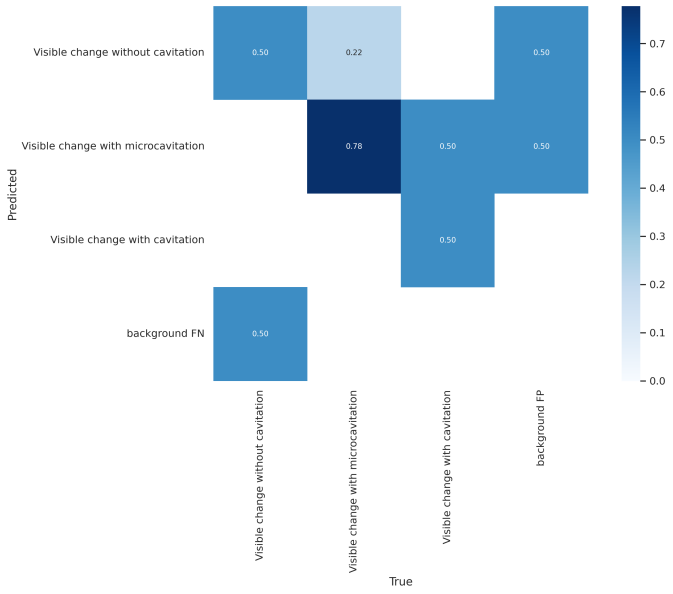

Fig. 4. Confusion matrix

In this confusion matrix, 50 percent of all tooth images are classified correctly as 'change without cavitation'. 78 percent of all tooth images are classified correctly as 'change without cavitation and 50 percent of all tooth images are classified correctly as 'change with cavitation'.

A precision curve, a recall curve and an F1-curve were also generated to estimate the performance of the model.

The precision curve determines how close or dispersed the measurements are to each other. The recall curve determines the percentage of a certain class correctly identified.

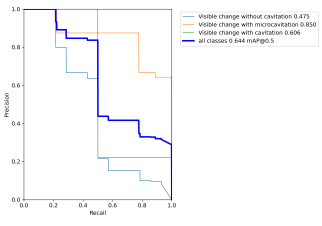

Fig. 5. PR curve

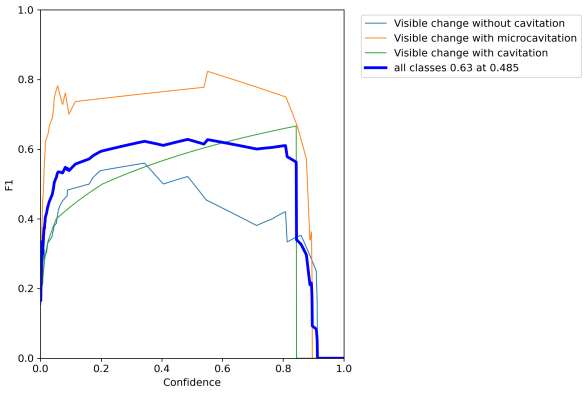

Fig. 6. F1 Curve

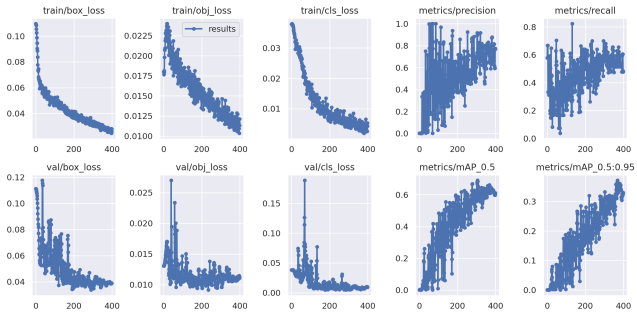

Fig. 7. Results

B. With Augmentation

To improve the performance of the model, several transformation techniques were used to enhance the performance of the model including blurring, contrast limited adaptive histogram equalization, gray coloring, changing brightness, changing contrast, and decreasing image compression.

After performing augmentation on the dataset, noticable improvement of the model's performance were noticed. We got 78% mean average precision(mAP) after augmenting the images, which was 64% when the model was trained without augmenting the images. This confusion matrix shows the overall performance of the model in this phase.

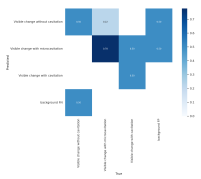

Fig. 8. Confusion matrix

After augmenting the images, 47 percent of all tooth images are classified correctly as 'change without cavitation'. 78 percent of all tooth images are classified correctly as 'change without cavitation and 50 percent of all tooth images are classified correctly as 'change with cavitation'.

A precision curve, a recall curve and an F1-curve were also generated to estimate the performance of the model after performing augmentation.

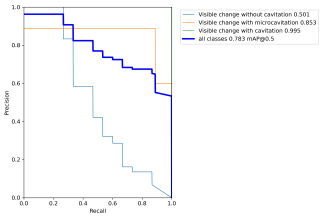

Fig. 9. PR curve

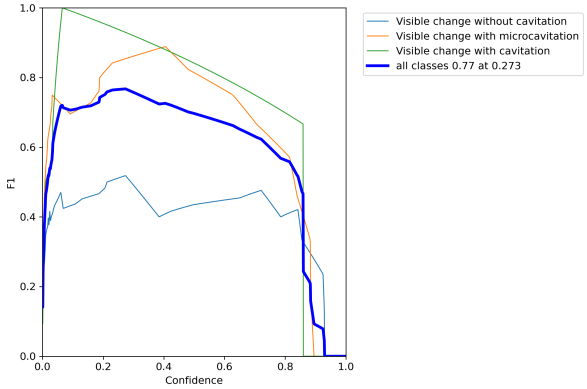

Fig. 10. F1 Curve

VI. Conclusion

According to the results of this in-vitro study of universal smartphone microphotography, photos of tooth decay captured by different smartphones have significant differences. When these photos were combined with a capable computer vision algorithm, an automated diagnostic model was created that was nearly as accurate in classifying tooth destruction as moderately skilled dental practitioners.

References

[1] M. B. Diniz, J. Rodrigues, and A. Lussi, "Traditional and novel caries detection methods," *Contemporary approach to dental caries*, pp. 105–128, 2012.

[2] S. Kositbowornchai, S. Siriteptawee, S. Plermkamon, S. Bureerat, and D. Chetchotsak, "An artificial neural network for detection of simulated dental caries," *International Journal of Computer Assisted Radiology and Surgery*, vol. 1, no. 2, pp. 91–96, 2006.

[3] R. Siva Kumar, "Identification of early caries in human tooth using histogram and power spectral analysis," *International Journal of Biomedical Engineering and Technology*, vol. 1, no. 4, pp. 465–472, 2008.

[4] E. D. Berdouses, G. D. Koutsouri, E. E. Tripoliti, G. K. Matsopoulos, C. J. Oulis, and D. I. Fotiadis, "A computer-aided automated methodology for the detection and classification of occlusal caries from photographic color images," *Computers in biology and medicine*, vol. 62, pp. 119–135, 2015.

[5] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[6] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in ct images," in *2015 12th conference on computer and robot vision*. IEEE, 2015, pp. 133–138.