THE DEEP LEARNING MODEL FOR DECAYED-MISSING-FILLED TEETH DETECTION: A COMPARISON BETWEEN YOLOV5 AND YOLOV8

Maya Fitria[1], Yasmina Elma[1], Maulisa Oktiana[1*], Khairun Saddami[1], Rizki Novita[2], Rizkika Putri[3], Handika Rahayu[1], Hafidh Habibie[1] and Subhan Janura[1]

(Received: 19-Mar.-2024, Revised: 17-May-2024, Accepted: 6-Jul.-2024)

ABSTRACT

Tooth decay is a dental condition characterized by the deterioration of tooth tissue originating from the outer surface and progressing to the pulp. Severe tooth decay, evolving into cavities, necessitates timely intervention to avert more serious dental-health issues. Common treatment procedures include filling and extraction of affected teeth. Presently, dentists conduct examinations for tooth decay by manually tallying affected, missing and filled teeth using an odontogram—a human tooth code diagram. This data is then recorded in patients' dental medical records. Recognizing the need for automation in assessing patients' experiences of tooth decay, this research endeavors to develop a model capable of detecting decayed, missing and filled teeth using variations of the YOLOv5 and YOLOv8 model architectures. The results of the training evaluation demonstrate the efficacy of YOLOv5l with a learning rate of 10^{-2}, exhibiting a high precision value of 0.97, a recall of 0.858 and a mean average precision (mAP) of 0.904 within 1 hour and 18 minutes. According to the curves obtained in the training process, YOLOv5l shows great performance on the dental caries dataset, but precautions like early stopping are needed for a reliable and generalizable model. In contrast, YOLOv8 offers better training stability and larger variants perform better on the dental caries dataset, improving detection capabilities with continued training epochs.

KEYWORDS

Caries detection, Detection model, Deep learning, DMF-T, Tooth decay.

1. INTRODUCTION

Although teeth are often known as the strongest part of the human body, they possess a vulnerable inner layer called the dental pulp tissues [1], [2], [3]. This tissue is vulnerable to bacteria and traumas that can lead to several tooth diseases [2]-[3]. One of the common chronic dental diseases is dental caries [4]. It is a complex infectious oral disease that progressively and accumulatively infects hard dental tissue, resulting in teeth loss [5]-[6]. The untreated caries teeth can cause pain and over an extended period, can cause inflammation to develop leading to subsequent swelling [7]. Numerous epidemiological and clinical studies additionally have suggested that tooth loss, particularly due to dental caries, could potentially be a risk indicator for cardiovascular and cognitive disorders [8]-[9]. Dental caries commonly occurs in children and almost 100% of adults [10]. Based on the Basic Health Research of Indonesia (RISKESDAS), 45,3% of the Indonesian population experience dental and oral health problems, with notable prevalence associated with cavities and damaged teeth and dental caries accounts for 88% of the severity prevalence [11]-[12].

Commonly, dentists employ manual examination methods to evaluate dental caries in each patient in the dental clinic or hospital, involving the inspection of cavity count (for teeth decay), the number of missing teeth and the count of filled teeth. Subsequently, dentists manually record the position of the teeth infected with caries in a form called the odontogram, an instrument to record the dental status of a person recorded in visual format [13]. This examination of caries status within the population typically requires the computation of the Decayed, Missing and Filled Teeth, known as the DMF-T index, to serve the preventive, curative and rehabilitative care, as well as for determination of dental-

1. M. Fitria, Y. Elma, M. Oktiana, K. Saddami, H. Rahayu, H. Habibie, and S. Janura are with Department of Electrical and Computer Eng., Universitas Syiah Kuala, Banda Aceh, Indonesia. Emails: mayafitria@usk.ac.id, yasmin06@mhs.usk.ac.id, maulisaoktiana@usk.ac.id, khairun.saddami@usk.ac.id, handika@mhs.usk.ac.id, habibie19@mhs.usk.ac.id, subhan.j@mhs.usk.ac.id
2. R. Novita is with Faculty of Dentistry, Universitas Syiah Kuala, Banda Aceh, Indonesia. Email: drg_rizkinovita@usk.ac.id
3. R. Putri is with Polyclinic of Dental and Oral Medicine, Regional General Hospital dr. Zainoel Abidin (RSUZA), Banda Aceh, Indonesia. Email: rizkikaputri@gmail.com

health status in a community [14]-[15]. As reported in interviews with dentists from Regional Hospital Zainoel Abidin, Aceh, Indonesia, conducting the dental caries assessment through DMF-T in a population is a time- consuming process. This is due to the assessment that requires a meticulous inspection of each decayed, missing and restored tooth individually, followed by the record count and manual calculation of the DMF-T index based on the gathered information. In epidemiological research, dental status such as DMF-T status, is used to describe caries prevalence in a certain population [16]. These records are utilized in forensics science, as intra-oral information is important to determine characteristics of individuals or corpses through criminal investigation or other civil cases [13], [16]. However, the odontogram of the patient is not completely filled by the dentists due to being preoccupied with treating other patients or the odontogram forms being run out [17]. Thus, an automated dental caries detection system is preferable to tackle this problem.

Several scientific studies have been undertaken to identify and detect dental caries, particularly employing the Convolutional Neural Network algorithm. A study conducted by Baydakar et al. utilized the U-Net and VGG-16 techniques to detect the cavities in radiographic bitewing images, resulting in 48% detection accuracy [18]. A similar study on radiographic bitewing dental data was also carried out by Kumari et al. employing an image enhancement process using CLAHE and FOC-KCC and a training process using M-ResNet-RNN. However, assessing dental caries in a population is not feasible and the utilization of radiographic images for this purpose is discouraged due to the potential risks associated with X-ray exposure [19]. To minimize the use of X-rays in the automated detection of dental caries, Fitria et al. undertook a study utilizing dental clinical images for the detection process using CNN architecture [20]. The work employed five sides of dental clinical dataset images; namely, anterior, left buccal, right buccal, upper occlusal and lower occlusal. The development model was conducted on 1400 augmented images implementing ResNet-50 architecture. However, the performance of this model is considered sub-optimal, as the datasets exhibit significant variability and the missing and filled teeth also need to be taken into account in addition to the classification of caries and non-caries cases [21]. In addition, the inclusion of other parts in the dataset images, such as gums, normal teeth, lips and various anatomical features creates a challenge for the system in accurately identifying the cavity areas, resulting in a relatively lower accuracy.

This research proposed a baseline work to address the limitation of manual caries inspection conducted by dentists. Moreover, this research aims to overcome the disadvantages of the CNN model developed in [20] by developing a deep-learning model to identify the caries experience based on Decayed, Missing, Filled Teeth (DMF-T) using a popular object-detection model, You Only Look Once (YOLO) [22]. This object-detection model is considered capable of detecting multiple objects in an image by using the bounding box technique for the object [23]. Decayed, missing and filled teeth are the objects that are detected in the work. Two different versions of YOLO models were implemented in this work, namely YOLOv5 and YOLOv8, as both of them tend to provide higher accuracy than other versions [22], [24]-[25]. Moreover, YOLOv5 is chosen to be adopted, as the models usually yield a significant accuracy with unaffected model's real-time performance [26]. Conversely, as the latest version of YOLO models, YOLOv8 is selected for its advancement, manifesting in a new neural-network architecture succeeding YOLOv5 [27]. The model features an anchor-free detection head that simplifies the detection process and improves accuracy. The clinical dataset used in this work is the dental clinical images obtained by Fitria et al [20]. The results of this work are expected to be a basis for future research to contribute to the practical development of dental diagnostic tools and telemedicine applications.

2. METHOD

Figure 1 shows the procedure conducted in this work, which involves four stages; 1) Problem analysis; 2) Dataset pre-processing; 3) Model development and verification; and 4) Result analysis. The procedure will be discussed in detail in the sub-sections below.

2.1 Dataset and Pre-processing

The datasets employed in this research were sourced from the dataset utilized by Fitria et al. in [20], gathered from the Dental and Oral Polyclinic of Regional Hospital Zainoel Abidin Banda Aceh, Indonesia. The dataset consists of 350 images identified in the caries class. However, only 294 of the caries images were selected, based on the considerations of light intensity, object clarity and image

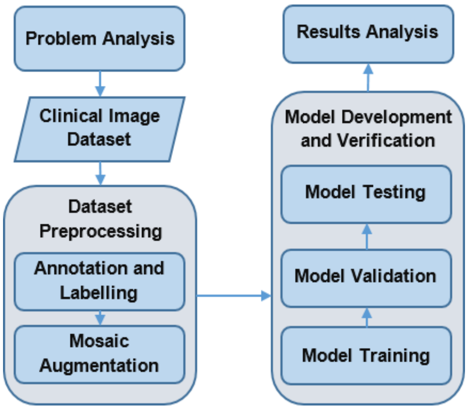


Figure 1. Research methodology.

sharpness, aiming to provide a strong feature representation. The removed dataset is considered weak in clarity, as some of the teeth are not exposed enough in the pictures and there is less light intensity and less sharpness, as the pictures contain more shadows, especially in the hindmost teeth. The dataset taken into account is discussed and chosen involving the dentists.

Figure 2 exhibits the sample of five sides of teeth used as datasets in this work; namely, labial (a), right buccal (b), left buccal (c), lower occlusal (d) and upper occlusal (e). Out of 294 images taken into consideration, there are 70 labial images, 58 right buccal images, 58 left buccal images, 52 lower occlusal images and 56 upper occlusal images. Moreover, the filled teeth shown in the images show temporary fillings of the patients.

(a) (b) (c) (d) (e)

Figure 2. A sample of five sides of teeth used as the dataset.

The pre-processing stage was conducted in this work to obtain optimized images for model training. The processes involved annotation, labeling and the mosaic augmentation process. The annotation and labeling process was carried out using the Roboflow image-labeling framework [28]. Dental image datasets were uploaded into Roboflow, where the annotation process involved placing bounding boxes on the images. The annotation process involved all authors of this work in collaboration with three dental experts and the annotated images were calibrated within the annotators. Subsequently, the annotated images were labeled according to their respective classes; decayed teeth were labeled as a red box, missing teeth were labeled as a yellow box and filled teeth were labeled as a blue box. For instance, there are two labels in the left image of Figure 3, decayed and missing teeth box, respectively and two boxes for decayed and filled teeth appear in the right image of Figure 3.

To enhance the diversity of the datasets and to make the model more robust, mosaic augmentation was established in the datasets. This type of augmentation involves the replacement and transformation of a certain area in the image with mosaic, such as breaking the images into small blocks [29]-[30]. This augmentation method is employed to assess the detection precision across various image dimensions that resemble the original image with a lower objective lens [31]. The reason for employing mosaic augmentation is to assess the detection precision across various image dimensions resembling those captured with a less powerful lens. Figure 4 illustrates the images that have already been augmented by implementing the mosaic augmentation technique. In total, there are 708 augmented images in the datasets, distributed among training, validation and testing data with a ratio of 88:8:4, respectively. Table 1 represents the distribution of data before and after augmentation and the allocation of training, validation and testing data.

Figure 3. Bounding-box technique and image annotation in the datasets.

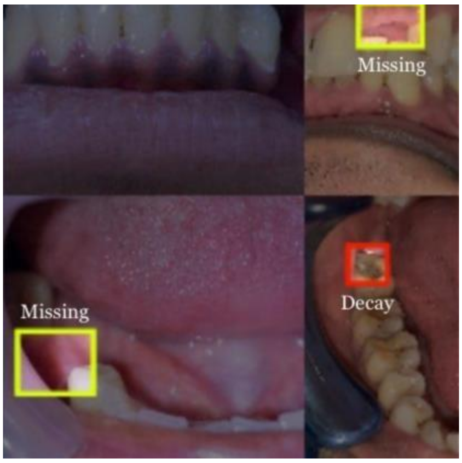


Figure 4. Mosaic augmentation.

Table 1. Distribution of datasets.

Dataset	**D**	**M**	**F**	**Total**	**Augmented Dataset**
Training	71	120	103	207	621
Validation	25	33	22	58	58
Testing	10	15	13	29	29
Total	106	168	148	294	708

2.2 Model Development and Verification

The model was developed to detect the caries condition; decayed, missing and filled teeth. The training process was carried out by implementing different variants of YOLOv5 and YOLOv8. The YOLOv5 network is divided into three parts; the backbone for the feature extraction on an input image using Cross-Stage Partial Network (CSPNet); the neck component for refining extracted features from the backbone; and the detection head for object detection using Feature Pyramid Network (FPN) [32]-[33]. YOLOv5 provides five different versions; namely, YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x, where every version provides different trade-offs in terms of calculation speed, average precision and the depth of channel [33]-[34]. Figure 5 shows the scales of corresponding versions of YOLOv5. The higher the versions, the larger and more accurate the models become, yet the slower the calculation speed [33]. Thus, this work selected the first versions only, which are YOLOv5n, YOLOv5s, YOLOv5m and YOLOv5l to avoid a long running time during the training process. Similar to YOLOv5, YOLOv8 also consists of a backbone network, a neck segment and a detection head [35]. However, YOLOv8 employs FPN for feature extraction in the backbone part and Cross-Layer Connection (CLC) in the neck [35]. The YOLOv8 versions specifically chosen in this experiment were YOLOv8n, YOLOv8s, YOLOv8m and YOLOv8l to prevent prolonged computation time. YOLOv8 incorporates features that make it a highly precise object detector, particularly through the use of an anchor-free detection head [36]-[37]. This approach simplifies the architecture of the model and improves its accuracy in predicting object locations. This enhancement is especially advantageous for datasets containing objects of various shapes and sizes.

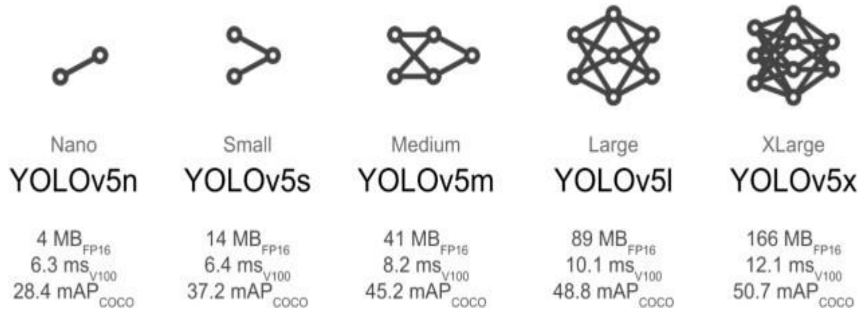


Figure 5. Variants of YOLOv5 [31].

The shift to an anchor-free detection head, away from the anchor box method used in previous YOLO versions, streamlines the detection process and boosts accuracy [36]. The variants of YOLOv8 can be seen in Figure 6. Considering the time constraints during the experiment, the hyper-parameters reported in this work, such as epoch, batch size, optimizer, momentum and learning rates, were set as shown in Table 2.

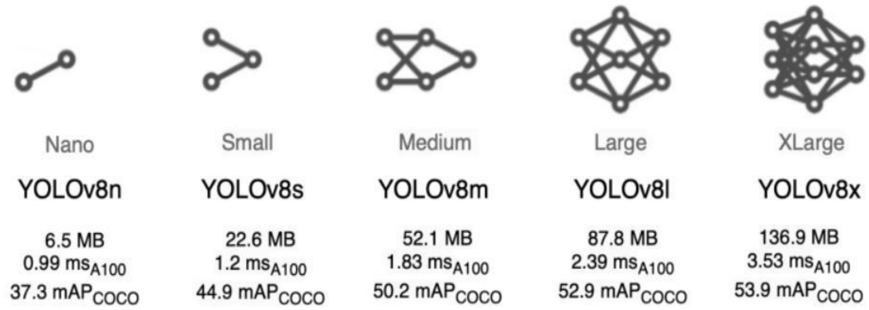


Figure 6. Variants of YOLOv8.

Table 2. Hyper-parameters for model development.

Hyper-parameters of YOLOv5 and YOLOv8		
Hyper-parameter	YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l	YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l
Epoch	400	100
Batch Size	16	16
Optimizer	SGD	SGD
Momentum	0.9	0.9
Learning Rate	10^{-2}, 10^{-3}, 10^{-4}	10^{-2}, 10^{-3}, 10^{-4}

The model-performance evaluation involves the analysis of the precision, recall and mean average precision (mAP) obtained by the models. Precision and recall serve as common metrics used for the evaluation of detection and classification models. Precision measures the accuracy of the model in identifying the positive class, while recall assesses how successfully the model identifies images of the positive class. In this work, precision is used to examine the accuracy of the model in identifying a dental caries image containing decayed, missing or filled teeth. Additionally, recall assesses the number of images containing caries that were correctly identified by the model.

The precision is computed by taking the ratio of true positive (TP) to the total predictions belonging to a positive class, as per Formula 1. On the other hand, recall is defined as the ratio of true positive (TP) to all predicted results, following Formula 2. True positive (TP) represents the number of accurately classified data as a positive class, while false positive (FP) denotes the number of incorrectly classified data as a positive class. In addition, true negative (TN) corresponds to the number of correctly classified data as a negative class. In contrast, false negative (FN) is the count of incorrectly classified data as a negative class. The precision and recall values increase with a higher TP count as well as with lower FP or FN values.

$$recall = \frac{TP}{TP+FN} \tag{1}$$

$$precision = \frac{TP}{TP+FP} \tag{2}$$

In the object-detection model, model output is not solely confined to the object class; it also includes additional outputs, such as bounding-box annotation for the detected object. Consequently, Mean Average Precision (mAP) is employed in this study, evaluating the average precision for decision values ranging from 0 to 1. The calculation of mAP, as depicted in Formula 3, involves N which represents the number of average precision (AP).

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{3}$$

Intersection over Union (IoU) assesses the overlap between the predicted bounding box and the ground truth bounding box. If the IoU exceeds 0.5, the value is considered True Positive; conversely, it is considered False Positive if the IoU is less than 0.5. To yield the Inter-precision for APi, recall maximum needs to be taken into account by using Formula 4.

$$AP_i = maxAP_i \tag{4}$$

Suppose that the automated system for detecting decayed, missing and filled teeth (DMFT) in dental clinical images is tested on a dataset containing 100 images and assume that there are 50 images with actual DMFT cases and 50 images without any DMF-T. Now, we consider the following hypothetical results:

- True positive (TP): The system correctly identifies 40 images with DMF-T.
- False positive (FP): The system incorrectly identifies 5 images without DMF-T as having DMF- T.
- False negative (FN): The system misses 10 images with actual DMF-T cases.

Using these results, now Precision, Recall and mAP are calculated as below:

- Precision = TP / (TP + FP) = 40 / (40 + 5) = 0.889.
- Recall = TP / (TP + FN) = 40 / (40 + 10) = 0.8.

To calculate the mAP, the average precision for each image in the dataset needs to be computed, which involves ranking the predicted DMF-T cases by their confidence scores. For the sake of brevity, it is assumed that there is an average precision of 0.85 for this example.

- Mean Average Precision (mAP) = Average precision across all images = 0.85.

In this example, our system achieved a precision of 0.889, indicating that 88.9% of the positive predictions were correct. The recall of 0.8 demonstrates that the system correctly identified 80% of the actual DMF-T cases. The mAP of 0.85 suggests that, on average, our model's predictions are highly accurate across all images in the dataset.

3. RESULTS AND DISCUSSION

3.1 Results

Tables 3 and 4 and Figures 7 and 8 show the training results of different versions of YOLOv5 and YOLOv8, respectively. It can be seen in Table 3 and Figure 7 that the smaller the learning rate tuned, the smaller the precision and recall obtained, resulting in a smaller mAP value of the model. Table 3 also indicates the increasing mAP value in every newer version of YOLOv5. However, the computation speed exhibited in Table 3 is inverse to the mAP value. The newer version of YOLOv5 employed, the longer the training time consumed. The mAP value yielded by the YOLOv5l version set with the learning rate of 10^{-2} is highlighted as the highest training result, outperforming the other versions with a mAP value of 90.4%, followed by YOLOv5s with a slightly different mAP value of 90.2%. Nevertheless, the calculation time of YOLOv5l tends to be longer than that of YOLOv5s, consuming one hour, 18 minutes and 42 seconds of training time, while YOLOv5s takes only 37 minutes and 28 seconds, which is the fastest running time amongst all models. A significant drop in mAP is also obtained in YOLOv5n, YOLOv5s and YOLOv5m tuned with a learning rate of 10^{-4}, where the mAP values are 36.9%, 47.5% and 56%, respectively.

Table 3. Result of YOLOv5.

Architecture	Learning Rate	Precision	Recall	mAP	Time
YOLOv5n	10^{-2}	0.954	0.803	0.878	45.24 m, s
	10^{-3}	0.831	0.712	0.772	47.46 m, s
	10^{-4}	0.483	0.313	0.369	45.36 m, s
YOLOv5s	10^{-2}	0.97	0.861	0.902	37.28 m, s
	10^{-3}	0.931	0.729	0.811	47.58 m, s
	10^{-4}	0.621	0.435	0.475	48.04 m, s
YOLOv5m	10^{-2}	0.971	0.856	0.895	40.05 m, s
	10^{-3}	0.90	0.785	0.841	01.04.12 h, m, s
	10^{-4}	0.788	0.485	0.56	01.00.24 h, m, s
YOLOv5l	10^{-2}	0.97	0.858	0.904	01.18.42 h, m, s
	10^{-3}	0.941	0.888	0.835	01.34.02 h, m, s
	10^{-4}	0.793	0.547	0.641	01.45.04 h, m, s

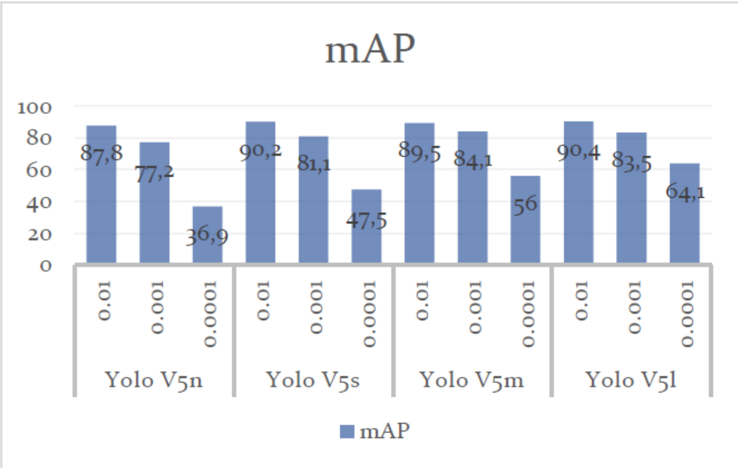


Figure 7. mAP value of different versions of YOLOv5 adjusted with different learning rates.

Similar to the YOLOv5, the four versions of YOLOv8 yielded a smaller mAP values as the learning rate decreases (Figure 8). Based on Tabel 4, the highest mAP value is received by YOLOv8m with a learning rate of 10^{-2} with a 90.6% mAP value, surpassing other models. This value is followed by YOLOv8l, YOLOv8n and YOLOv8s, yielding mAP values of 89.7%, 87.8% and 87.1%.

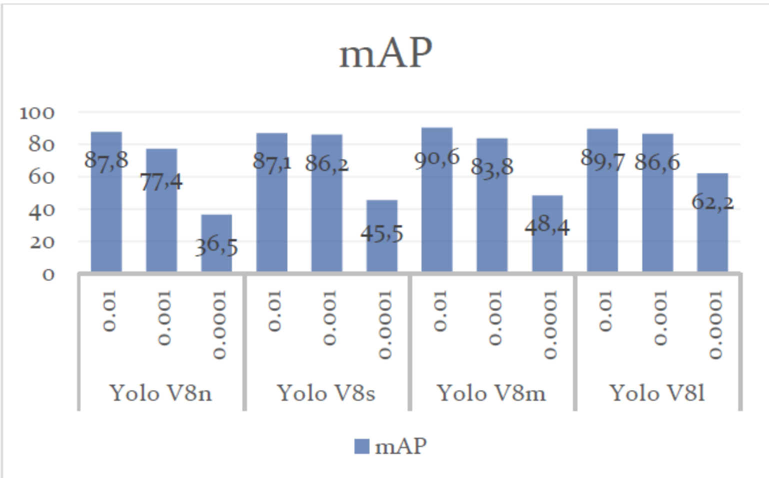


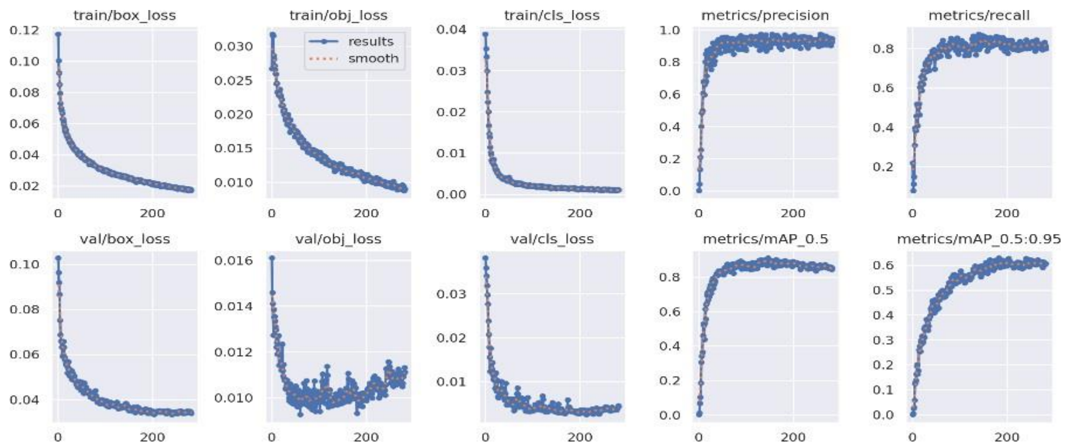
Figure 8. mAP value of different versions of YOLOv8 adjusted with different learning rates.

A notable decrease of mAP value also appears in every variant of the YOLOv8 set with a learning rate of 10^{-4}. Despite the similarity in the behavior in both YOLOv5 and YOLOv8, the training time of YOLOv8 tends to be longer than that of YOLOv5, with the fastest time of 4 hours, 15 minutes and 18 seconds and the slowest time of 5 hours, 32 minutes and 52 seconds, while the fastest computation time of YOLOv5 is 37 minutes and 28 seconds by YOLOv5s and the slowest time of one hour, 45 minutes and 4 seconds by YOLOv5l.

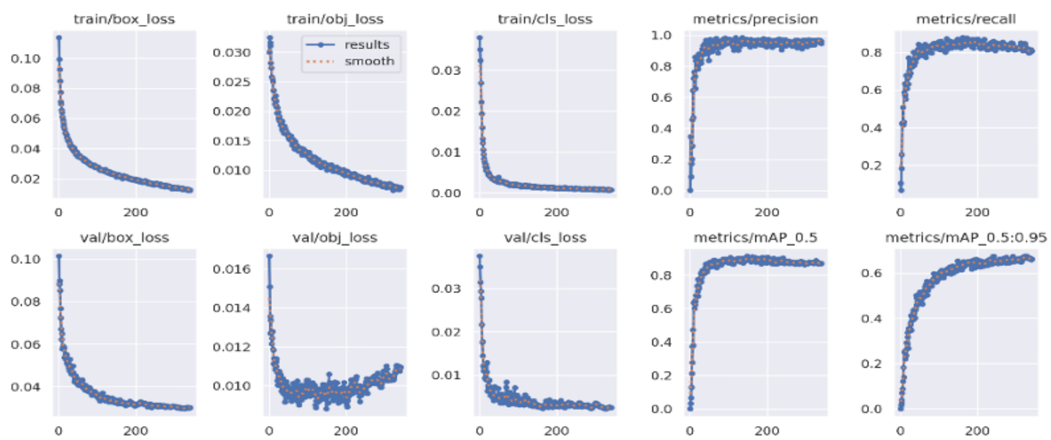
Table 4. Results of YOLOv8.

Architecture	Learning Rate	Precision	Recall	mAP	Time
YOLOv8n	10^{-2}	0.954	0.803	0.878	04.15.18 h, m, s
	10^{-3}	0.946	0.685	0.774	04.19.08 h, m, s
	10^{-4}	0.427	0.417	0.365	04.20.34 h, m, s
YOLOv8s	10^{-2}	0.955	0.811	0.871	04.55.06 h, m, s
	10^{-3}	0.956	0.778	0.862	04.22.32 h, m, s
	10^{-4}	0.573	0.454	0.454	04.25.28 h, m, s
YOLOv8m	10^{-2}	0.954	0.846	0.906	04.57.30 h, m, s
	10^{-3}	0.945	0.79	0.838	05.02.04 h, m, s
	10^{-4}	0.668	0.457	0.484	05.06.06 h, m, s
YOLOv8l	10^{-2}	0.953	0.839	0.897	05.20.16 h, m, s
	10^{-3}	0.986	0.806	0.866	05.21.04 h, m, s
	10^{-4}	0.751	0.518	0.622	05.32.52 h, m, s

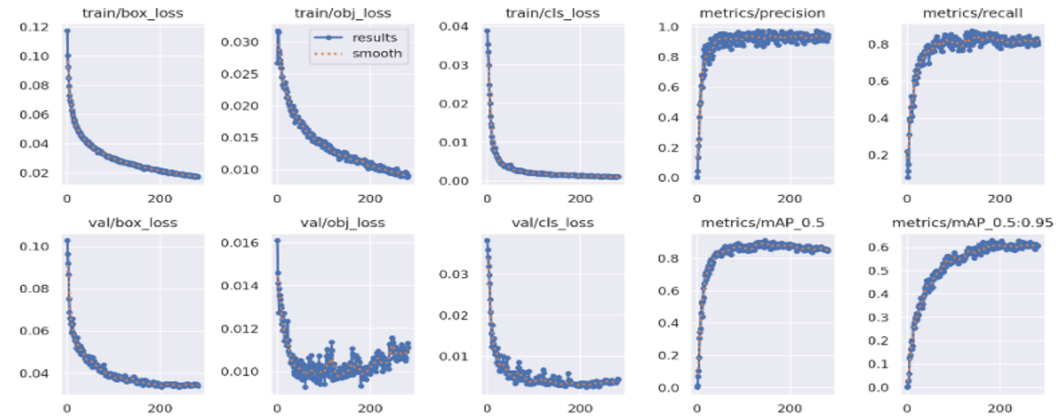
Figure 9 presents a comparison of training results across various versions of YOLOv5 on a dental caries dataset. It's evident from the graph that training with YOLOv5 produces the most effective model. The training curve highlights the outstanding performance of YOLOv5l. Based on Figure 10, YOLOv8 presents a different aspect compared to YOLOv5. YOLOv8 generates a more stable training graph than YOLOv5.



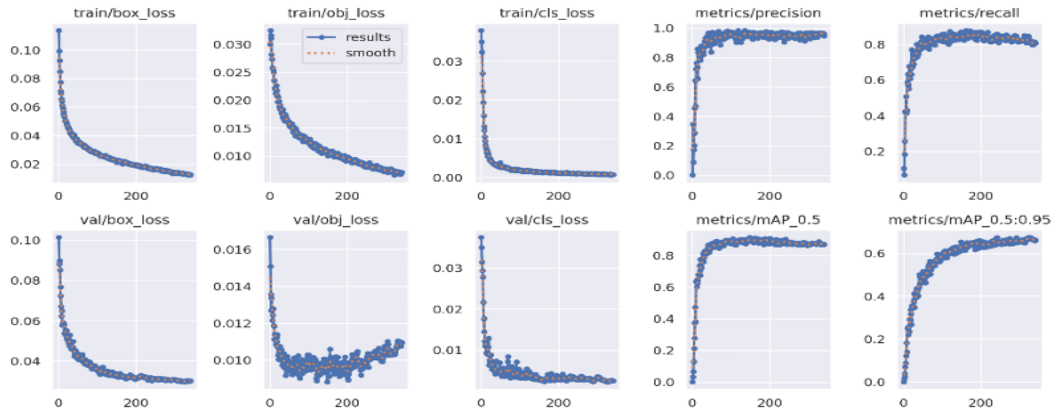
(a) YOLOv5n



(b) YOLOv5

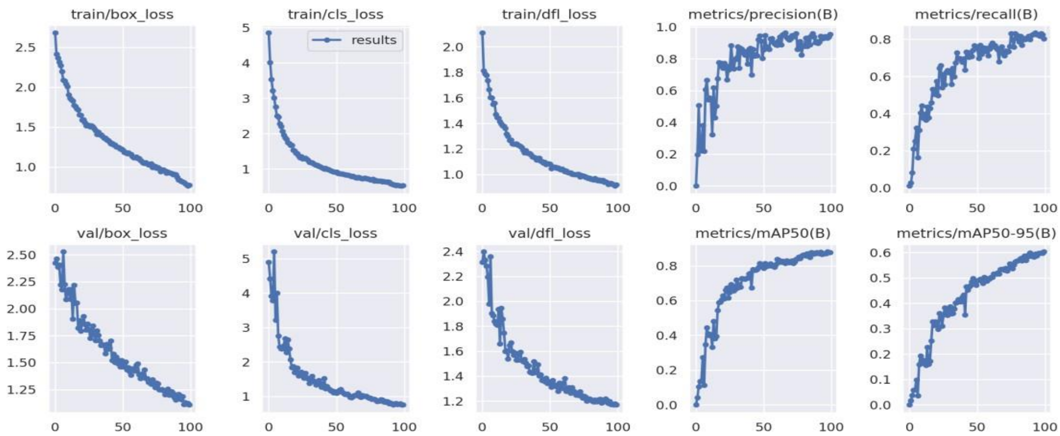


(c) YOLOv5m

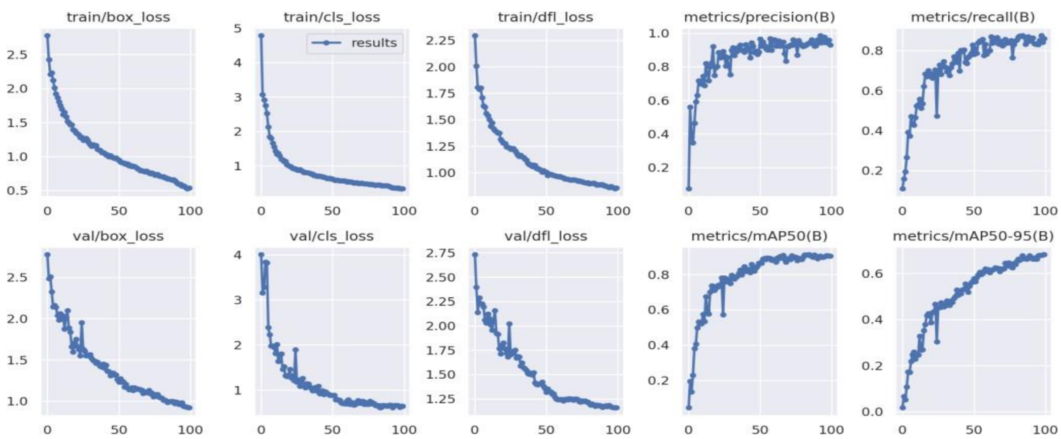


(d) YOLOv5l

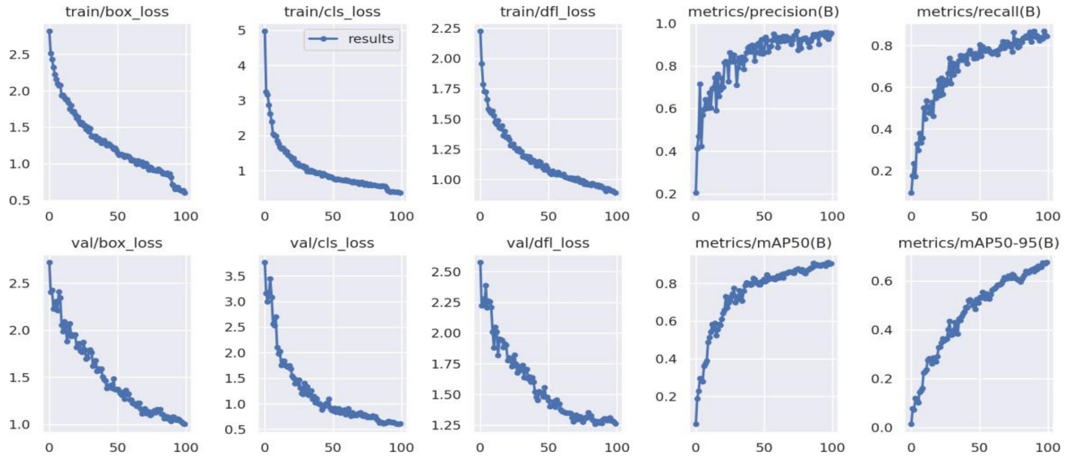
Figure 9. Training curves of YOLOv5.



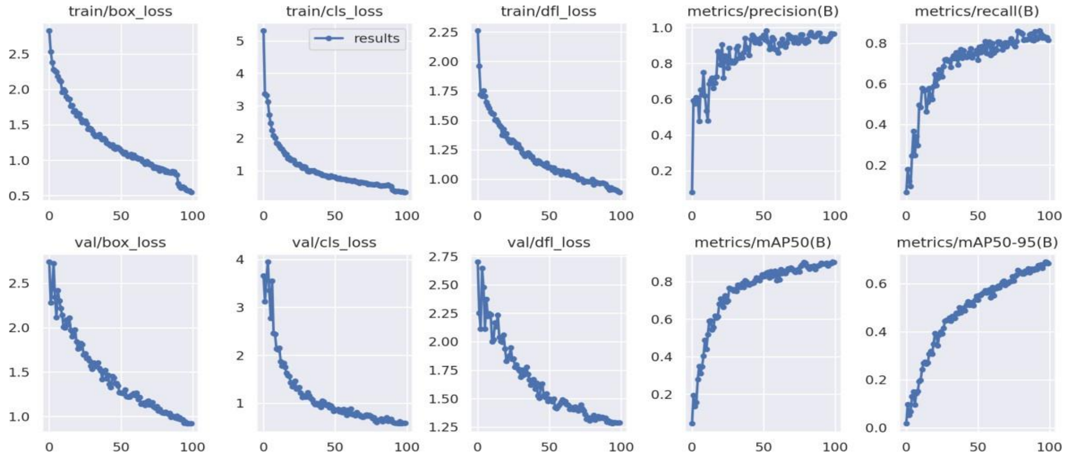
(a) YOLOv8n



(b) YOLOv8s



(c) YOLOv8m



(d) YOLOv8l

Figure 10. Training curves of YOLOv8.

Figures 11 and 12 depict the model detection of YOLOv5 and YOLOv8 variants on a sample dental clinical image, adjusted with a learning rate of 10^{-2}. It is shown that YOLOv5l detects the caries indications better than YOLOv8s with accuracies above 90% detected as decayed teeth (labeled as Karies) and filled teeth (labeled as Tambal). On the other hand, there is a filled tooth on the image detected by YOLO8s yielding an accuracy of 51%.

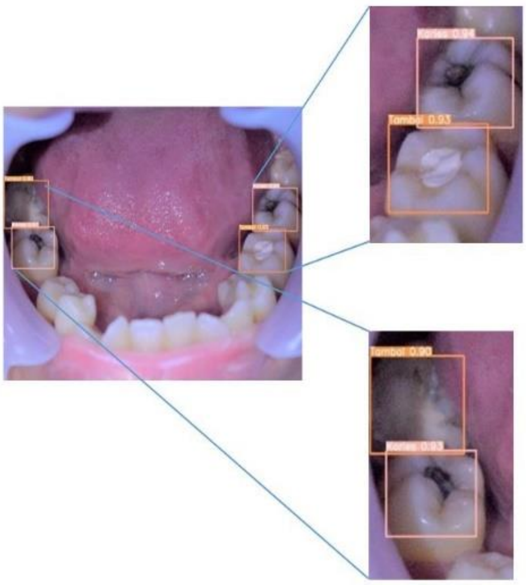


Figure 11. Model detection results of YOLOv5l with the learning rate of 10^{-2}.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 03, September 2024.

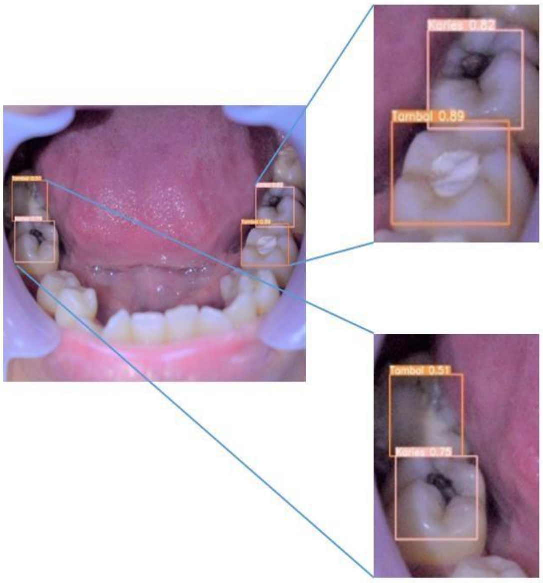


Figure 12. Model detection results of YOLOv8m with the learning rate of 10^{-2}.

Figure 13 visualizes the confusion matrix of the YOLOv5l model tested on 29 dental images. It can be seen that 87% of the missing teeth are correctly detected, while 77% of the decayed class were correctly detected and 93% of the filled teeth were correctly detected. This indicates that the model yielded robust performance and is effectively identifying the correct DMF-T in practical scenarios.

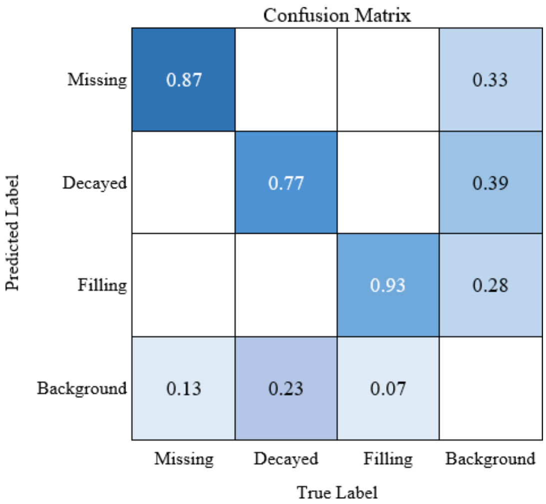


Figure 13. Confusion matrix of YOLOv5l with a learning rate of 10^{-2}.

3.2 Discussion

According to Tables 3 and 4, it is shown that the differences in mAP values and training times can indeed impact the practical applicability of the models in real-world scenarios. Higher mAP values generally indicate better model performance, which can lead to more accurate detection of decayed, missing and filled teeth (DMF-T) in dental clinical images. However, models with higher mAP values may also have longer training times and require more computational resources, which can be a concern in resource- limited settings or when deploying the model on edge devices. On the other hand, models with faster training times and lower computational requirements may have slightly lower mAP values, but may be more suitable for real-world applications, where efficiency and resource constraints are critical factors. In such cases, the trade-off between model accuracy and computation time should be carefully considered based on the specific requirements of the target environment and application. In this study, we have aimed to strike a balance between model accuracy and computation time by selecting the YOLOv5 and YOLOv8 versions that offer a reasonable compromise between these factors. However, it is acknowledged that the optimal choice may vary depending on the specific use case and available resources.

Based on Figure 9, YOLOv5l yielded the most outstanding training curve. However, validation results for this model indicate overfitting occurring after surpassing 100 epochs, a trend also observed with the YOLOv5 medium model. Despite this, YOLOv5l still stands out as the superior option overall, although it necessitates the implementation of early stopping in order to prevent overfitting. Early stopping can prevent validation loss from increasing by halting the training process if the validation loss stops decreasing or starts to increase, which is a sign of overfitting. Overfitting, a common issue in machine learning, occurs when a model learns noise or irrelevant patterns from the training data, thereby hindering its ability to generalize well to unseen data. Early stopping serves as a regularization technique to mitigate overfitting by halting the training process when the model's performance on a validation dataset starts to decline. By doing so, it ensures that the model maintains its ability to generalize effectively.

Compared to YOLOv5, YOLOv8 in Figure 10 generated more stable training curves. The larger the size of YOLOv8, the better the model learns about the dental caries dataset. The same trend is also observed in the number of training iterations. The more epochs utilized, the closer the model gets to convergence. It's evident that YOLOv8 large yields the best-performing model. The stability of the training graph in YOLOv8 suggests improved training dynamics and possibly better handling of the dataset's complexities. Additionally, the correlation between model size and performance indicates that larger YOLOv8 variants are more adept at capturing intricate features within the dental caries dataset. Moreover, the convergence of the model with increased epochs implies a continuous improvement in learning, leading to enhanced model accuracy.

In summary, while YOLOv5l demonstrates exceptional performance on the dental caries dataset, precautions such as early stopping are necessary to ensure the model's reliability and generalization capabilities. This approach would lead to the development of a robust model capable of accurately detecting dental caries in real-world applications. However, YOLOv8 showcases superior stability in training, with larger variants demonstrating enhanced performance on the dental caries dataset. The convergence of the model with more training epochs signifies the ongoing refinement of the model's understanding, ultimately resulting in improved detection capabilities.YOLOv8 provided new features that improved its detection capabilities, increasing both accuracy and efficiency. This model particularly excels in segmentation tasks, offering precise segmentation and classification of various image parts, making it highly effective for diverse applications, like medical imaging and autonomous vehicle navigation.

4. CONCLUSION

This research proposed a comparison of deep learning-based models for dental caries detection, characterized by decayed, missing and filled teeth. Two different object detection architectures are implemented in this work; YOLOv5 and YOLOv8, including their variants. The models' performances are compared by calculating their precision, recall and mAP values. The results show that YOLOv5l and YOLOv8m outperformed other variants with the mAP values of 90.4% and 90.6%, respectively. However, the computational time required by YOLOv8m is considered extensive; namely, around four hours 57 minutes 30 seconds, while YOLOv5s takes only one hour, 18 minutes and 42 seconds. The YOLOv8 annotation format is unique and requires precise detailing of objects in images, usually using bounding boxes and labels. Preparing a dataset for YOLOv8 involves annotating numerous images, which can be a time-consuming task. The quality and accuracy of these annotations directly influence the model's ability to learn and make precise predictions. To further enhance performance, the YOLOv8 model should be integrated into the existing image-enhancement process.

ACKNOWLEDGMENTS

This research was supported by Lembaga Penelitian dan Pengabdian Masyarakat Universitas Syiah Kuala. In addition, we thank the dentists and nurses from Polyclinic of Dental and Oral Medicine, Regional General Hospital dr. Zainoel Abidin (RSUZA), who helped and assisted the authors throughout the process of dataset-gathering.

REFERENCES

[1] S. Prabhu, A. B. Acharya and M. V. Muddapur, "Are Teeth Useful in Estimating Stature?," Journal of Forensic and Legal Medicine, vol. 20, no. 5, pp. 460–464, DOI: 10.1016/j.jflm.2013.02.004, 2013.

[2] G. T. Huang, "Pulp and Dentin Tissue Engineering and Regeneration: Current Progress," Regenerative Medicine, vol. 4, no. 5, pp. 697–707, DOI: 10.2217/rme.09.45, 2009.

[3] X. Huang et al., "Microenvironment Influences Odontogenic Mesenchymal Stem Cells Mediated Dental Pulp Regeneration," Frontiers in Physiology, vol. 12, p. 656588, 2021.

[4] L. Cheng et al., "Expert Consensus on Dental Caries Management," Int. Journal of Oral Science, vol. 14, no. 1, p. 17, 2022.

[5] H. Nilsson, J. Sanmartin Berglund and S. Renvert, "Longitudinal Evaluation of Periodontitis and Tooth Loss among Older Adults," Journal of Clinical Periodontology, vol. 46, no. 10, pp. 1041–1049, 2019.

[6] M. J. Y. Yon, S. S. Gao, K. J. Chen, D. Duangthip, E. C. M. Lo and C. H. Chu, "Medical Model in Caries Management," Dentistry Journal, vol. 7, no. 2, p. 37, 2019.

[7] T. Kikuiri, K. Saito, A. Iida, Y. Yoshimura, Y. Yawaka and T. Shirakawa, "Occurrence of Subcutaneous Emphysema during a Caries Filling Procedure: A Case Report," Pediatric Dental Journal, vol. 32, no. 3, pp. 211–215, 2022.

[8] M. Zhou, J. Dong, L. Zha and Y. Liao, "Causal Association between Periodontal Diseases and Cardiovascular Diseases," Genes, vol. 13, no. 1, p. 13, 2021.

[9] 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, 2018.

[10] M. Rathee and A. Sapra, "Dental Caries," [Online], Available: https://www.ncbi.nlm.nih.gov/books/NBK51699/, 2019.

[11] Balitbangkes, "National Report of Basic Health Research 2018 (in Indonesian)," Indonesian Ministry of Health, p. 206, 2019. [Online], Available: https://repository.badankebijakan.kemkes.go.id/id/eprint/3514/1/Laporan%20Riskesdas%202018%20Nasional.pdf, [Accessed: February 2024].

[12] N. D. Ardiyanti, R. Adhani and I. Hatta, "Correlation between DMF-T Caries Index, Consumption of Drinking Water and Tooth Brushing Behavior in Indonesian Communities (in Indonesian)," Dentin, vol. 6, no. 1, 2022.

[13] E. S. Wardhana, S. Suryono, A. Hernawan and L. E., Nugroho, "Evaluation of Format and Security of Dental Electronic Medical Record Systems in General Hospital Based on Legislation," Odonto: Dental Journal, vol. 9, Special Issue 1, pp. 80-89, 2022.

[14] G. Moradi, A. M. Bolbanabad, A. Moinafshar, H. Adabi, M. Sharafi and B. Zareie, "Evaluation of Oral Health Status Based on the Decayed, Missing and Filled Teeth (DMFT) Index," Iranian Journal of Public Health, vol. 48, no. 11, p. 2050, 2019.

[15] A. Alami, S. Erfanpoor, E. Lael-Monfared, A. Ramezani and A. Jafari, "Investigation of Dental Caries Prevalence, Decayed, Missing and Filled Teeth (DMF-T and DMF-T Indices) and the Associated Factors among 9-11 Years Old Children," Research Square, pp. 1-18, DOI: 10.21203/rs.2.21545/v1, 2020.

[16] J. A. Daza-Cardona, J. Vargas-Ramírez and M. A. Guapacha-Sánchez, "Doing Odontograms and Dentists in the Classroom. Materiality and Affect in Dental Education," Tapuya: Latin American Science, Technology and Society, vol. 4, no. 1, p. 1968635, 2021.

[17] E. D. Fadhillah et al., "Smart Odontogram: Dental Diagnosis of Patients Using Deep Learning," Proc. of 2021 IEEE Int. Electronics Symposium (IES), pp. 532-537, DOI: 10.1109/IES53407.2021.9594027, Surabaya, Indonesia, 2021.

[18] I. S. Bayrakdar et al., "Deep-learning Approach for Caries Detection and Segmentation on Dental Bitewing Radiographs," Oral Radiology, vol. 38, no. 4, pp. 468-479, pp. 1–12, 2021.

[19] Y. Al-Hadeethi, M. Sayyed, H. Mohammed and L. Rimondini, "X-ray Photons Attenuation Characteristics for Two Tellurite Based Glass Systems at Dental Diagnostic Energies," Ceramics International, vol. 46, no. 1, pp. 251–257, 2020.

[20] M. Fitria et al., "Development of Intraoral Clinical Image Dataset for Deep Learning Caries Detection," Proc. of the 2023 IEEE 2nd Int. Conf. on Computer System, Information Technology and Electrical Engineering (COSITE), pp. 194–198, DOI: 10.1109/COSITE60233.2023.10249428, Banda Aceh, Indonesia, 2023.

[21] L. Que et al., "Prevalence of Dental Caries in the First Permanent Molar and Associated Risk Factors among Sixth-grade Students in São Tomé Island," BMC Oral Health, vol. 21, no. 1, pp. 1–10, 2021.

[22] K. Kim, K. Kim and S. Jeong, "Application of YOLO v5 and v8 for Recognition of Safety Risk Factors at Construction Sites," Sustainability, vol. 15, no. 20, p. 15179, 2023

[23] T. Diwan, G. Anirudh and J. V. Tembhurne, "Object Detection Using YOLO: Challenges, Architectural Successors, Datasets and Applications," Multimedia Tools and Applications, vol. 82, no. 6, pp. 9243–9275, 2023.

[24] V. D. Matta et al., "Single Use Plastic Bottle Recognition and Classification Using Yolo V5 and V8 Architectures," Proc. of the Int. Conf. on Cognitive Computing and Cyber Physical Systems, Part of the Book Series: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 537, pp. 99– 106, Springer, 2023.

[25] E. Casas, L. Ramos, E. Bendek and F. Rivas-Echeverría, "Assessing the Effectiveness of YOLO Architectures for Smoke and Wildfire Detection," IEEE Access, vol. 11, pp. 96554 – 96583, DOI: 10.1109/ACCESS.2023.3312217, 2023.

[26] J. Terven, D.-M. Córdova-Esparza and J.-A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-nas," Machine Learning and Knowledge Extraction, vol. 5, no. 4, pp. 1680–1716, 2023.

[27] I. P. Sary, S. Andromeda and E. U. Armin, "Performance Comparison of YOLOv5 and YOLOv8 Architectures in Human Detection using Aerial Images," Ultima Computing: Jurnal Sistem Komputer, vol. 15, no. 1, pp. 8–13, 2023.

[28] Q. Lin, G. Ye, J. Wang and H. Liu, "RoboFlow: A Data-centric Workflow Management System for Developing AI-enhanced Robots," Proc. of the 5th Conf. on Robot Learning, vol. 164, pp. 1789–1794, [Online], Available: https://proceedings.mlr.press/v164/lin22c.html, PMLR, 2022.

[29] Y. Lee, J. Choi and K. Jo, "VSNet: Vehicle State Classification for Drone Image with Mosaic Augmentation and Soft-label Assignment," Proc. of the Asian Conference on Intelligent Information and Database Systems, Part of the Book Series: Lecture Notes in Computer Science, vol. 13995, pp. 109–120, 2023.

[30] F. Dadboud, V. Patel, V. Mehta, M. Bolic and I. Mantegh, "Single-stage UAV Detection and Classification with YOLOv5: Mosaic Data Augmentation and PANet," Proc. of the 2021 17th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–8, DOI: 10.1109/AVSS52988.2021.9663841, Washington, USA, 2021.

[31] T. B. Pun, A. Neupane, R. Koech and K. Walsh, "Detection and Counting of Root-knot Nematodes Using YOLO Models with Mosaic Augmentation," Biosensors and Bioelectronics: X, vol. 15, p.100407, DOI: 10.1016/j.biosx.2023.100407, 2023.

[32] Q. Song et al., "Object Detection Method for Grasping Robot Based on Improved YOLOv5," Micro-machines, vol. 12, no. 11, p. 1273, 2021.

[33] L. Wang et al., "Investigation into Recognition Algorithm of Helmet Violation Based on YOLOv5-CBAM- DCN," IEEE Access, vol. 10, pp. 60622–60632, 2022.

[34] W. Sheng et al., "Symmetry-based Fusion Algorithm for Bone Age Detection with YOLOv5 and ResNet34," Symmetry, vol. 15, no. 7, p. 1377, 2023.

[35] B. Selcuk and T. Serif, "A Comparison of YOLOv5 and YOLOv8 in the Context of Mobile UI Detection," Proc. of the Int. Conf. on Mobile Web and Intelligent Information Systems, Part of the Book Series: Lecture Notes in Computer Science, vol. 13977, pp. 161– 174, Springer, 2023.

[36] G. Wen, M. Li, Y. Luo, C. Shi and Y. Tan, "The Improved YOLOv8 Algorithm Based on EMSPConv and SPE-head Modules," Multimedia Tools and Applications, vol. 83, pp. 61007–61023, DOI: 10.1007/s11042-023-17957-4, 2024.

[37] H. Wang, C. Liu, Y. Cai, L. Chen and Y. Li, "YOLOv8-QSD: An Improved Small Object Detection Algorithm for Autonomous Vehicles Based on YOLOv8," IEEE Transactions on Instrumentation and Measurement, vol. 73, DOI: 10.1109/TIM.2024.3379090, 2024.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 03, September 2024.

ملخص البحث:

تآكل الأسنان وضع صحّي يصيب الأسنان ويتميّز بتآكل نسيج الأسنان، حيث يبدأ من السّطح الخارجي للسّنّ ويتطوّر إلى أن يصل إلى لُبّ السّنّ. ويتطلّب التّآكل الشّديد للأسنان تدخّلاً في الوقت المناسب للحيلولة دون قضايا أشدّ خطراً. وتشمل إجراءات العلاج الشّائعة الحشو والاستئصال للأسنان المتضرّرة والمفقودة والمحشوّة باستخدام مخطّط رموز يفصّل الحالة السّنّية لكلّ مريض. ويتمّ تسجيل المعلومات ذات العلاقة في السّجلّات الطّبية للمرضى.

يسعى هذا البحث إلى تطوير نموذجٍ قادرٍ على الكشْف عن الأسنان المتآكلة والمفقودة والمحشوّة باستخدام البنى المعروفة باسم (يولو 5) و (يولو 8) بأشكالها المختلفة. وقد بيّنت النّتائج فعالية النّموذج (يولو 5)/البنية (l) بمعدل تعلّم مقداره (10^{-2})، محققاً دقّةً عاليةً في الكشْف بلغت 97%، ومعدّل استرجاعٍ قدرُهُ (0.858)، ومتوسط دقة كشفٍ قدره (0.904) في غضون زمنٍ مقداره ساعة واحدة و 18 دقيقة.

وبناءً على المنحنيات التي تمّ الحصول عليها في عملية التّدريب، حقّق النّموذج يولو 5/ البنية (l) أداءً عالياً عند تطبيقه على مجموعة بيانات تآكل الأسنان، لكن مع احتياطاتٍ مثل التّوقّف المبكّر ليكون النّموذج أكثر موثوقية وقابلية للتّعميم . بالمقابل، يوفّر نموذج يولو 8 استقرارية تدريب أفضل، وتعمل الأشكال الأكبر منه بشكل أفضل على مجموعة بيانات تآكل الأسنان.