

Multi-granularity tooth analysis via YOLO-based object detection models for effective tooth detection and classification

Samah AbuSalim¹, Nordin Zakaria¹, Aarish Maqsood², Abdul Saboor¹, Kwang Hooi Yew¹,
Norehan Mokhtar³, Said Jadid Abdulkadir¹

¹Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Perak, Malaysia

²Department of Agricultural Engineering, Bahauddin Zakariya University, Multan, Pakistan

³Dental Simulation and Virtual Learning Research Excellence Consortium, Department of Dental Science, Advanced Medical and Dental Institute, Universiti Sains Malaysia, Penang, Malaysia

Article Info

Article history:

Received Aug 11, 2023

Revised Oct 12, 2023

Accepted Dec 2, 2023

Keywords:

Deep learning
Dental informatics
Intra-oral image
Tooth detection
You only look once

ABSTRACT

Effective and intelligent methods to classify medical images, especially in dentistry, can assist in building automated intra-oral healthcare systems. Accurate detection and classification of teeth is the first step in this direction. However, the same class of teeth exhibits significant variations in surface appearance. Moreover, the complex geometrical structure poses challenges in learning discriminative features among the tooth classes. Due to these complex features, tooth classification is one of the challenging research domains in deep learning. To address the aforementioned issues, the presented study proposes discriminative local feature extraction at different granular levels using you only look once (YOLO) models. However, this necessitates a granular intra-oral image dataset. To facilitate this requirement, a dataset at three granular levels (two, four, and seven teeth classes) is developed. YOLOv5, YOLOv6, and YOLOv7 models were trained using 2,790 images. The results indicate superior performance of YOLOv6 for two-class classification achieving a mean average precision (mAP) value of 94%. However, as the granularity level is increased, the performance of YOLO models decreases. For, four and seven-class classification problems, the highest mAP value of 87% and 79% was achieved by YOLOv5 respectively. The results indicate that different levels of granularity play an important role in tooth detection and classification.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Samah AbuSalim

Department of Computer and Information Sciences, Universiti Teknologi PETRONAS

Seri Iskandar 32610, Perak, Malaysia

Email: samah_21000332@utp.edu.my

1. INTRODUCTION

Deep learning, a subset of artificial intelligence (AI), has recently emerged as a transformative technology. With its near-to-human-level image classification accuracy [1], it finds its applications in various fields such as astronomy [2], food [3], and healthcare [4]. Deep learning models have proven the ability to analyze and extract intricate features from large medical image datasets and subsequently have gained considerable attention from researchers in dental studies [4], [5]. These models are primarily employed for tooth classification, disease diagnosis, treatment planning, and oral health risk assessment.

Accurate tooth detection and classification are important for an automated intra-oral treatment system. Such systems provide dentists with numerous advantages. This includes but is not limited to the identification of a region of interest, decreased diagnostic error, improved prognostic ability, and formulation

of a treatment plan [6]. Additionally, time efficiency and low cost of such systems are other valuable benefits.

Extensive studies have been undertaken to explore the implementation of deep learning-based object detection methods for tooth detection and classification. These studies have focused on improvements in models such as faster region-based convolutional neural network (R-CNN) [7], you only look once (YOLO) [8], and AlexNet [9]. Various dental datasets including radiographic, cone beam computed tomography (CBCT), and intra-oral images are employed to train and test these models [10]. Among these datasets, Bitewing, periapical, and panoramic image datasets are most commonly used. However, these images have inherent limitations. This includes the presence of tooth ghost images, low resolution, reduced contrast, overlaps, and angulation [11]. All these unwanted features introduce noise into data. CBCT on the other hand, offers high-quality three-dimensional volumetric information. By adopting this approach, problems associated with distortion and superimposition of bony and dental structures are alleviated. However, CBCT also introduces specific challenges, including noisy images, unclear tooth edges, and the occurrence of artifacts such as human skulls [12].

In recent times, intra-oral dental images have emerged as an alternative solution to the aforementioned issues. They have gained popularity in tooth disease diagnosis studies. Intra-oral images yield valuable insights into a patient's oral health condition and the simplicity of solution is an added benefit. These images offer several advantages, including i) eliminating the need for specialized equipment for data acquisition, ii) providing rich features despite small image sizes, and iii) requiring low computational resources for image processing and tooth detection tasks. On the other hand, issues such as partial occlusion of the tooth, overlapping, rich geometrical structures, and variations in illumination [13] present challenges in the accurate identification and detection of individual teeth.

To address the aforementioned issues, the presented study hypothesizes that the discriminative local features of intra-oral images are naturally buried in the images' varied granularity patches. Studies have shown that object detection accuracy improves when objects are analyzed at several scales or levels of complexity. However, little studies has been done to determine the role of granularity in enhancing intra-oral image-based tooth detection and classification [10].

For this purpose, the presented study investigates the effect of granularity on tooth detection and classification using intra-oral images. Firstly, various granular levels for teeth are formulated. This includes two, four, and seven classes granularity levels. Based on this formulation, a dataset was previously proposed and developed [10]. The dataset is named granular intra-oral image dataset (GIOI). Secondly, three variants of YOLO models namely YOLOv5, YOLOv6, and YOLOv7 are trained and tested on each granular level. The mAP values for YOLO models at each level indicates enhanced performance as compared to state-of-the-art models named faster R-CNN-50, faster R-CNN-101 and faster R-CNN-152. The results are then compared, and a conclusion is drawn. The following are the contributions of the presented study:

- Evaluating the effect of systematic reduction in granularity on the performance improvement of deep learning models for tooth detection and classification.
- Establishing a conclusive ground to gauge the effectiveness of granularity in improving tooth detection and classification accuracy using intra-oral images.

The study is divided into the following sections. Section 2 reviews literature related to deep learning models-based tooth detection and classification and establishes the need for the presented study. Section 3 describes the methods and materials used to investigate the problem. Section 4 discussed the results and formulates a conclusion. Finally, section 5 presents the conclusion and future work.

2. RELATED WORK

Within the realm of dental informatics, numerous methodologies have been devised to facilitate dental diagnosis, employing various types of radiographic images like bitewing, periapical, panoramic images, and CBCT images. Consequently, different researchers have conducted diverse investigations into tooth detection and classification, utilizing different data formats. In recent years, deep convolutional neural networks (CNNs) have shown great potential in tooth detection and numbering tasks, yielding promising results. Various types of neural networks, including visual geometry group-16 (VGG16), residual network (ResNet), YOLO, and faster R-CNN, have been developed specifically for object detection and have been successfully adopted in dental research. The following subsection provides a review of related work on the topic of deep learning models-based tooth detection and classification, as well as their limitations.

By employing 1250 periapical images, Chen *et al.* [14] employed faster R-CNN features for tooth detection and numbering. To improve upon the benchmark faster R-CNN, they devised three post-processing techniques based on prior knowledge. The achieved accuracy for tooth type determination ranged between 71.5% and 91.7%. Additionally, Tuzoff *et al.* [15] proposed a two-stage system, where faster R-CNN was

employed for tooth detection, followed by a VGG-16 network for numbering maxillary and mandibular teeth in a single image. To evaluate the system's performance, a dataset comprising 1,574 anonymized panoramic radiographs was utilized for tooth detection and numbering, following the Fédération Dentaire Internationale (FDI) notation. The study reported that the CNN-based system demonstrated a performance level comparable to radiologists, achieving a sensitivity of 0.9941 and a precision of 0.9945. Yasa *et al.* [16] conducted an analysis of 1,125 bitewing images using a faster R-CNN approach to detect and number teeth. They employed a pre-trained GoogLeNet Inception v2 faster R-CNN network for preprocessing the training datasets through transfer learning. The proposed neural network achieved a tooth numbering precision of 0.9293. Similarly, Lee *et al.* [17] introduced R-CNN method for tooth segmentation, utilizing individual annotations on 30 panoramic radiographs of adults. To mitigate overfitting, an augmentation technique was employed, while the tooth structures were detected and localized through a fine-tuning process using a fully deep learning method based on the mask R-CNN model. They achieved an F1 score of 0.875, precision of 0.858, recall of 0.893, and a mean intersection over union (IoU) of 0.877 for automated tooth segmentation.

Kaya *et al.* [8] evaluated the efficiency of a deep learning system in automating tooth detection and numbering on pediatric panoramic radiographs. They utilized YOLOv4, an advanced object detection model, and used a dataset comprising 4545 pediatric panoramic X-ray images. The reported evaluation metrics include a mean average precision (mAP) value of 92.22%, a mean average recall (mAR) value of 94.44%, and a weighted-F1 score of 0.91. An AI-driven clinical dentistry decision-support system was developed by Yilmaz *et al.* [18] using deep-learning techniques, leading to a reduction in diagnostic interpretation errors and time. A study compared two alternative deep learning approaches for tooth classification in dental panoramic radiography, namely faster R-CNN and YOLO-V4. The YOLO-V4 approach demonstrated superior performance in terms of tooth prediction accuracy, detection speed, and its ability to identify impacted and erupted third molars achieved an average precision of 0.9990, a sensitivity of 0.9918, and an F1 score of 0.9954. On the other hand, the faster R-CNN method obtained an average precision of 0.9367, a sensitivity of 0.9079, and an F1 score of 0.9221. Du *et al.* [19] introduced a teeth-detection approach specifically designed for processing CBCT images. The method proposed by the authors involved several key steps to achieve accurate tooth detection. The proposed method demonstrated a significant reduction in training and prediction times, achieving an 80% decrease in training time and a 62% decrease in prediction time compared to the faster R-CNN approach. Furthermore, Gerhardt *et al.* [20] introduced AI-driven tool that successfully accomplished the automated tasks of tooth detection, segmentation, and labeling, eliminating the need for manual intervention. The tool demonstrated remarkable performance with a general accuracy of 99.7% and 99% for tooth detection and labeling, including missing teeth, for both fully dentate and partially dentate patients, respectively. Miki *et al.* [21] introduced a tooth type classification CNN model based on the AlexNet network. They performed manual cropping of each tooth from a single 2D slice of CBCT images and then inputted the cropped 2D image to the network for tooth type classification, achieving a classification accuracy of 0.89.

Celik [22] evaluated the potential usefulness and accuracy of different network architectures in detecting impacted teeth including faster R-CNN with ResNet50, AlexNet, and VGG16 as backbones, and YOLOv3. A dataset of 440 panoramic radiographs from 300 patients was randomly divided for evaluation. The faster R-CNN technique achieved a mAP of 0.91 with ResNet50 as the backbone. VGG16 and AlexNet showed slightly lower performances with mAP values of 0.87 and 0.86, respectively. On the other hand, the YOLOv3 technique exhibited the highest detection efficacy, achieving a mAP of 0.96. The recall and precision values for YOLOv3 were 0.93 and 0.88, respectively. A fully convolutional network (FCN) based on GoogleNet for tooth detection was introduced by Muramatsu *et al.* [23]. Subsequently, tooth classification based on their types, including incisors, canines, premolars, and molars, was performed using a pre-trained ResNet-50 network. The researchers utilized a dataset of 100 dental panoramic radiographs to train and evaluate their object detection network, employing a 4-fold cross-validation method. The tooth detection sensitivity achieved was 96.4%, indicating a high capability in accurately detecting teeth. The overall accuracy of tooth detection was reported as 93.2%.

Table 1 summarize that CNN models, like faster R-CNN, YOLO, and AlexNet, have been utilized for tooth detection and classification. These models are trained on diverse dental image datasets which consist of X-Rays and CBCT images. The review identified that frequently class imbalance is present in these datasets, with certain tooth types being underrepresented. Moreover, variation in tooth shape, size, and position, presents various challenges in detection and classification task. Despite its ease of use, intra-oral images are rarely used in dental studies. Furthermore, it is identified that object detection models are sensitive to missing features such as broken or missing tooth. Similarly, variation in illumination, degree of occlusion and sensor induced noise affects the performance of these models. The importance of granularity in object detection is well established. It precisely localizes and recognize objects of various sizes and levels of visibility. The literature indicates that the effect of granularity in dental studies has seldom researched. This presents an open gap for further investigation.

Table 1. Summarize the reviewed studies for tooth detection using deep learning models

Authors name	Model	Dataset	Conclusion
Chen <i>et al.</i> [14]	Faster R-CNN	Periapical films	Faster R-CNN demonstrates precise localization of teeth with a high IoU value. However, the model proves sensitive to objects with missing features caused by broken teeth.
Tuzoff <i>et al.</i> [15]		Panoramic radiographs	The system failed to accurately identify a tooth adjacent to a missing tooth. Also, the model misclassified the partially occluded tooth in the background, such as the molar.
Yasa <i>et al.</i> [16]		Bitewing radiographs	The model correctly identified and numbered teeth but it has severe limitations in recognizing exact object contour.
Kaya <i>et al.</i> [8]	YOLOv4	Panoramic X-ray	The model yielded high mAP (92.22%) for small objects; however, the study has limited generalization ability.
Yilmaz <i>et al.</i> [18]	Faster R-CNN, YOLOv4	Panoramic radiography	In terms of tooth prediction accuracy and computational efficiency, the YOLOv4 model surpasses the faster R-CNN method.
Du <i>et al.</i> [19]	YOLOv3	CBCT	The model outperformed faster R-CNN in identifying teeth in the presence of high noise and “missing tooth”. However, errors are commonly encountered during the detection process, particularly in the following scenarios: severe malocclusion and extreme metal artifacts.
Miki <i>et al.</i> [21]	AlexNet	CBCT	The study achieves high classification accuracies however the study used sliced tooth images and small test data to produce the results. The model may underperform in complex dental images.
Celik [22]	Faster RCNN, YOLOv3	Panoramic radiographs	The YOLOv3 produced higher mAP (0.96) for tooth detection, however, the applicability of this study is limited to teeth with no or little occlusion.
Muramatsu <i>et al.</i> [23]	ResNet	Panoramic radiographs	Significant tooth detection sensitivity and classification accuracies were achieved. However, the model misclassified teeth which are at the background and under low illumination.

3. METHOD

In this section, the methodology employed to accomplish the primary objective of this study is presented. The overview of the model architecture and algorithm to detect teeth at multi-granularity levels using deep learning is presented in Figure 1. The detection pipelines are shown in four essential steps: dataset development, models development, experimental design, model evaluation matrices, and result analysis. The following subsections briefly describe each step and technique used in this work.

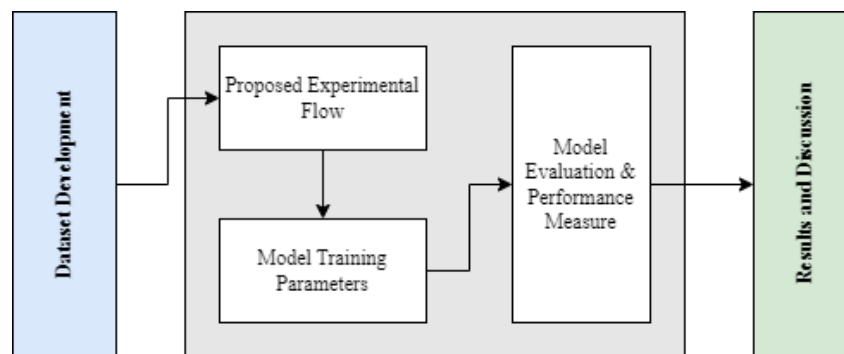


Figure 1. Research methodology flow

3.1. Dataset development

The dataset utilized in presented study comprises 2709 images and includes three granular levels. These levels are named two, four, and seven granular levels (i.e., 2CGL, 4CGL, and 7CGL respectively). The source images were collected from the Advanced Medical and Dental Institute at Universiti Sains Malaysia. The images were acquired from different dentists, captured at various times, distances, and lighting conditions. The dataset includes images of individuals from diverse genders and age groups ranging from 18 to 50, ensuring a diverse collection that encompasses a wide range of variations. The process of dataset development is explained in detailed in a previously presented study [10].

3.2. Proposed experimental flowchart

In this research study, three granularity levels, namely 2CGL, 4CGL, and 7CGL, are defined to represent varying levels of tooth detail. Each level corresponds to a different number of classes. Various models for object detection and classification emerge as the field evolves, however, YOLO's strengths in speed, simplicity, and generalization remain compelling options for various object detection challenges [18]. For this reasons, three YOLO models, YOLOv5, YOLOv6, and YOLOv7, are selected and configured with the appropriate number of classes for each granularity level. For each model, hyperparameter tuning is conducted using grid search. The models are trained and tested on GIOI at each granular level. Figure 2 shows the experiment's flowchart and the specific implementation described in the next section.

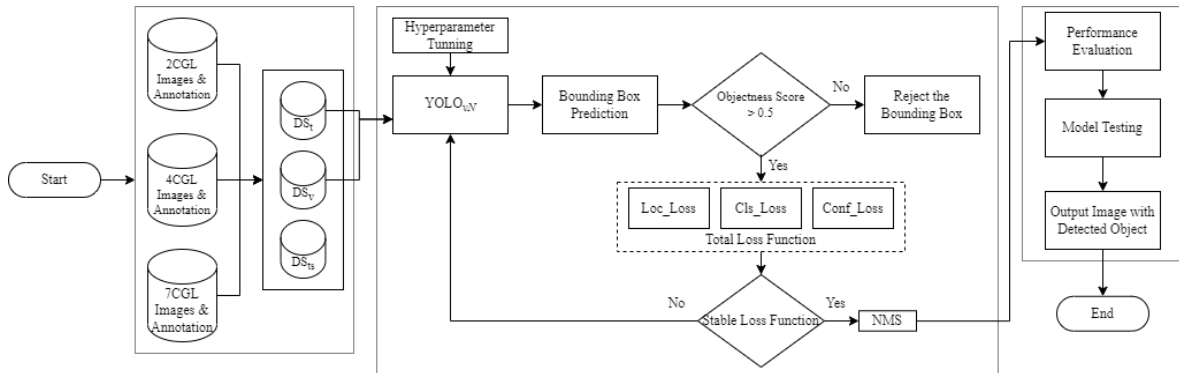


Figure 2. Experimental flowchart

3.3. YOLO model

YOLO is a state-of-the-art object detection and classification model. It consists of backbone, neck, and head modules. The backbone module predicts multiple bounding boxes (usually 2 or more) for each grid cell of the generated feature map. Associated confidence scores and class probabilities are also calculated. The objectness score, which is a part of the bounding box prediction, represents the confidence value for an object in the grid cell. It is a single scalar value ranging from 0 to 1, where 1 indicates high confidence (i.e., the bounding box contains an object) and 0 indicates low confidence (i.e., the bounding box does not contain an object). The presented study sets a threshold of 0.5 and above to accept a bounding box.

The model's loss function is a combination of several components designed to handle both object detection and classification tasks. It consists of three main parts which are explained as follow:

- Localization loss: measures the accuracy of the predicted bounding box coordinates compared to the ground truth bounding box coordinates.
- Classification loss: measures the accuracy of the predicted class probabilities compared to the ground truth class labels.
- Confidence loss: measures the accuracy of the objectness score compared to the ground truth objectness label, indicating whether an object is present in the bounding box or not.

The overall loss is the sum of these individual losses. During training, the model optimizes its parameters to minimize this combined loss, ensuring accurate bounding box predictions and class probabilities. Once a prediction is made, the model generates multiple bounding boxes for the same object as multiple grid cells may contain the same object. To consolidate and refine the detections, non-maximum suppression is applied as a post-processing step. Non-maximum suppression removes duplicate and overlapping bounding boxes, retaining only the most confident and non-overlapping detections. As a result, model detects and classifies an object in the image with improved accuracy.

3.4. Model training parameters

For our proposed dataset, to enable comprehensive evaluation of the dataset and to set a baseline mAP matrix, an emphasis is placed on state-of-the-art deep learning-based image classification models. A total of three models were initially chosen: YOLOv5, YOLOv6, and YOLOv7. Each of the deep learning models is trained, validated, and tested on the proposed dataset using the standard model training parameters described in Table 2.

Table 2. Standard model training parameters

SN	Training parameter	Value
1	Optimizer	Stochastic gradient decent (SGD)
2	Initial learning rate	0.01
3	Max epochs	100
4	Mini batch size	8
5	Execution environment	Graphics processing unit (GPU)

3.5. Model evaluation and performance measures

In this section, various performance measures and metrics utilized in this research are explained. These metrics have been derived from a "confusion matrix," which includes false negatives (FN), true negatives (TN), true positives (TP), and false positives (FP). Precision is a measure of the accuracy of detected objects. It represents the ratio of true positives (correctly detected objects) to the total number of detected objects (1). Recall is a measure of how well the algorithm detects all instances of the object. It represents the ratio of true positives to the total number of objects in the ground truth (2). F1-score is a measure of the balance between precision and recall (3). It is calculated as the harmonic mean of precision and recall. The average AP across multiple object categories is called mAP (4).

$$Precision = \frac{(TP)}{(TP + FP)} \quad (1)$$

$$Recall = \frac{(TP)}{(TP + FN)} \quad (2)$$

$$F1 - Score = \frac{(2TP)}{(2TP + FP + FN)} \quad (3)$$

$$mAP = \frac{1}{|classes|} \sum_{c \in classes} \frac{|TP|}{|TP| + |FP|} \quad (4)$$

4. RESULTS AND DISCUSSION

YOLO is an advanced object detection algorithm renowned for its real-time capabilities and end-to-end design [24]. Unlike traditional methods, YOLO performs object detection in a single forward pass through a CNN, swiftly identifying multiple objects within an image [25]. This efficiency has made YOLO a prominent choice in various applications within the field of computer vision. In this context, this section evaluated and presents the experimental results for three models of YOLO named; YOLOv5, YOLOv6, and YOLOv7 at three different levels of intra-oral images i.e., two classes granularity level (2CGL), four classes granularity level (4CGL), and seven classes granularity level (7CGL).

4.1. Two classes granularity level

This level consists of images that are classified as lower and upper jaws. These two classes have features that pose challenges in tooth detection. For example, object instances in at this level include the occlusion of the lower jaw by the upper jaw, and in some cases the occlusion significantly distorts the features of the lower jaw. Similarly, the degree of illumination within the class varies dramatically from the center of the image to the extreme right and left. This also affects the contrast level, leading to a gradual gradient shift in tooth color.

The performance of all three models observed in this study produced a dynamic range of results. For example, YOLOv5 results report an overall mAP of 0.89 as shown in Figure 3. The model has its strength in non-maximum suppression, which helps detect objects of varying sizes in tested images. This can also be seen in the model's recall values for both classes, which turn out to be 1 as shown in Figure 4(a). However, one of the strengths of CSPDarknet53 is local feature extraction, which may limit its ability to capture global contextual information [25]. The effect of this limitation is observed in the results. As YOLOv5's produced a lower average precision value (i.e., 0.88) for the class that is occluded. To improve a model's performance, its ability to perceive global features is important.

Contrary to YOLOv5, YOLOv6 gains its performance improvements from network elements such as VGG as the backbone. VGG networks have proven their ability to deduce finer features due to their smaller receptive fields. This plays a key role in improved accuracy for occluded object detection [26]. For this reason, the results show model's ability to optimally identify occluded classes with a higher average precision of 0.94. The model generated the highest value for mAP as well, i.e., 0.94 as shown in Figure 3. Additionally, this model achieves highest F1-score among other models for both classes as shown in

Figure 4(b). The results indicate the model’s superiority in larger object detection tasks where occlusion, varying size, and illumination are prominent object features.

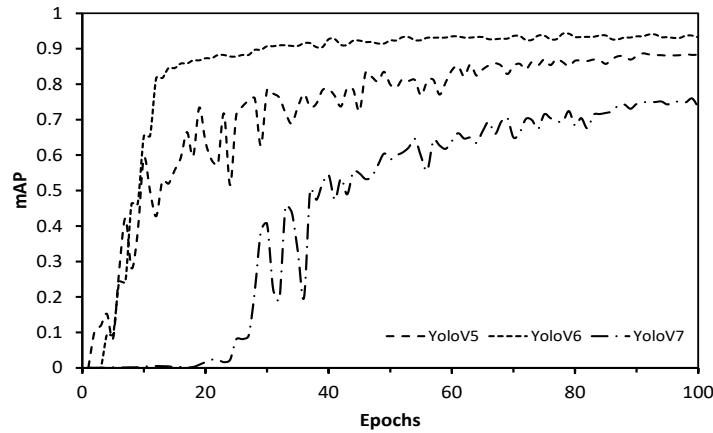


Figure 3. 2CGL mAP for YOLO models

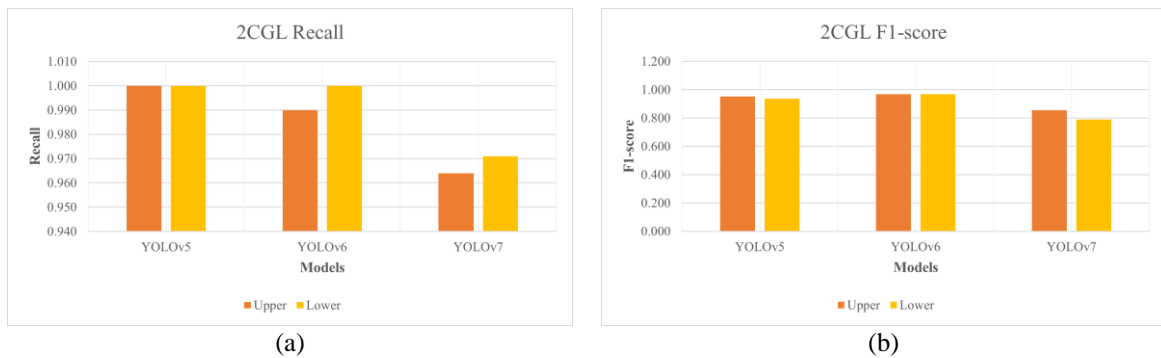


Figure 4. Results summary for 2CGL using YOLO models including (a) 2CGL recall and (b) 2CGL F1-score

The recent YOLO version, i.e., YOLOv7, is based on a deeper network design [27]. To improve the accuracy of object detection, the model proposes the trainable bag-of-freebies method. However, this approach combines problem-specific techniques and components to produce optimal mAP. We observed that, for two-class tooth detection, the model does not perform well on the dental dataset. The overall mAP of the model remains low at 0.72. This is a decrease of 23.4% in performance when compared to YOLOv6. Furthermore, as shown in Figures 4(a) and (b), the recall and F1-score achieved by YOLOv7 are lower than other models. The model also suffered when it came to correctly identifying occluded teeth. For example, the average precision for the upper and lower classes remains as low as 0.77 and 0.66, respectively. Overall, it is reported that YOLOv5 is the optimal model for the level 1 tooth granularity level. Finally, as discussed and presented in Table 3 and Figures 3, 4(a)-(b), it is identified that YOLOv6 is the optimal model for the 2-class granular level. In contrast to the findings detailed in a prior research paper [10] that examined faster R-CNN, our investigation revealed that YOLOv6 achieved the highest mAP of 0.94 for 2CGL, whereas faster R-CNN-101 yielded the lowest mAP of 0.84.

Table 3. 2CGL mAP, recall, and F1 scores for YOLO models

2CGL	mAP			Recall			F1-Score		
	YOLOv5	YOLOv6	YOLOv7	YOLOv5	YOLOv6	YOLOv7	YOLOv5	YOLOv6	YOLOv7
Upper	0.910	0.948	0.767	1.000	0.990	0.964	0.953	0.969	0.854
Lower	0.881	0.937	0.664	1.000	1.000	0.971	0.937	0.967	0.789

4.2. Four classes granularity level

This level offers finer granularity and groups teeth into four classes. Since the teeth are generally tightly coupled, at this level the issue of overlapping objects is present. Further, in the case of mandibular

teeth, artefacts such as shadows suddenly affect the illumination levels. Additionally, the tooth classes exhibit variations in size. However, within a class, the group of teeth has shape and size similarity if the tooth structure is centered in the middle of the image.

The object scale variation results in a deviation from the anchor box size, making it a challenging case for object detection. With the highest mAP of 0.87 as depicted in Figure 5, YOLOv5 outperformed the other two models. This indicates that tooth object clustering remains generally unaffected by scale variation. Additionally, precise labelling also plays a pivotal role in achieving optimal results.

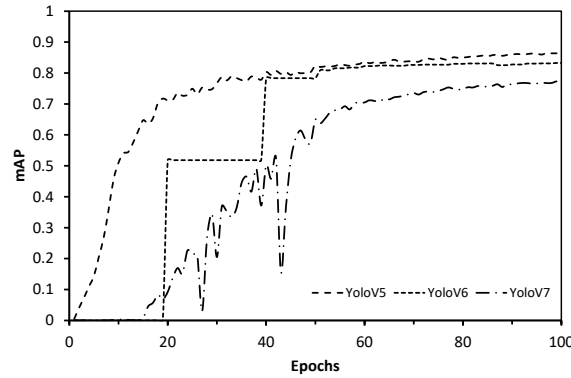


Figure 5. 4CGL mAP for YOLO models

The molar class is located at the back of the jaw. It is subject to low illumination and features that are not apparent. Among all models, YOLOv5 produced the highest average precision of 0.77 for the molar class as shown in Table 4. This shows that the multi-scale fusion technique used in YOLOv5 is optimally identifying objects with low illumination and fewer features. Furthermore, YOLOv5 achieves the highest recall values for all classes as depicted in Figure 6(a).

Despite its high performance in 2CGL (i.e., a two-class problem), the YOLOv6 model produced a relatively low mAP of 0.83 for 4CGL (i.e., a four-class problem) as shown in Figure 5. The highest average precision of 0.94 was observed for the incisor class; however, with an F1 score of 0.79, the model suffered greatly when it came to detecting the molar class as shown in Figure 6(b). With a mAP of 0.77, YOLOv7’s performance remains the lowest performing model among all others. For the most feature-enriched class, i.e., Incisor, the model’s F1 score was 0.94, which in the case of YOLOv5 was 0.97 as shown in Table 4 and Figure 6(b). Similarly, for the molar class, the model performed poorly as well and produced an average precision of 0.61 and an F1 score of 0.73. These results highlight that YOLOv5’s superiority over faster R-CNN at this level, as faster R-CNN-152 yielding the lowest mAP of 0.62, as referenced in [10].

Table 4. 4CGL mAP, recall, and F1 scores for YOLO models

4CGL Classes	mAP			Recall			F1		
	YOLOv5	YOLOv6	YOLOv7	YOLOv5	YOLOv6	YOLOv7	YOLOv5	YOLOv6	YOLOv7
Incisor	0.943	0.942	0.893	1.000	0.990	1.000	0.971	0.965	0.943
Canine	0.880	0.860	0.828	0.999	0.990	0.987	0.936	0.920	0.901
Premolar	0.866	0.836	0.775	1.000	0.980	0.988	0.928	0.902	0.869
Molar	0.778	0.692	0.608	0.993	0.940	0.900	0.872	0.797	0.726

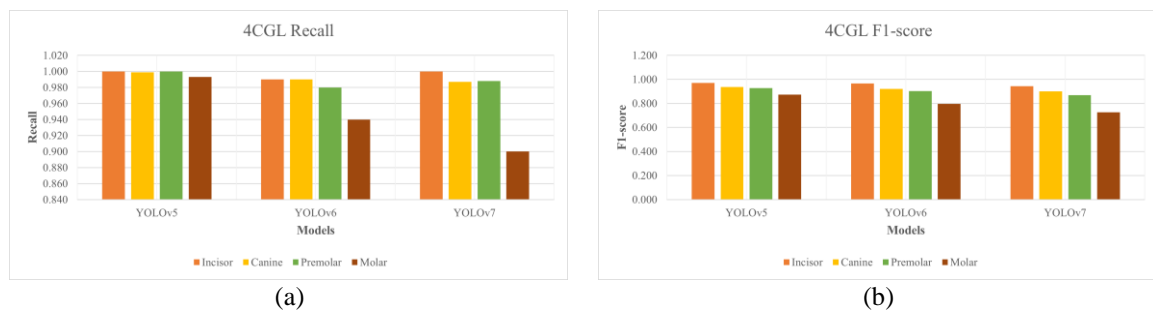


Figure 6. Results summary for 4CGL using YOLO models including (a) 4CGL recall and (b) 4CGL F1-score

4.3. Seven classes granularity level

The 7CGL is the finest granular level analyzed in this study. Each tooth is individually targeted. The feature similarity among classes is high as compared to other granular levels. Contrary to 2CGL and 4CGL, the tooth objects are subject to less occlusion, and illumination levels change gradually. With these attributes, the hypothesis is that at level three, object detection may yield higher accuracy.

However, as depicted in Figure 7, with the mAP of 0.80 achieved by the top performing model YOLOv5, this assumption is found to be false as higher mAPs are observed at other granular levels. Additionally, it is observed that as the object size becomes smaller, the average precision and recall values of all models decrease. For example, in the case of the 2nd Molar, which usually has the smallest pixel representation and lowest illumination levels in an intra-oral image, YOLOv7 resulted in the lowest values for mAP and recall, i.e., 0.44 and 0.47, respectively. For the same tooth class, the mAP and recall values yielded by YOLOv5 were significantly higher, i.e., 0.66 and 0.95, respectively as shown in Figure 7 and Figure 8(a).

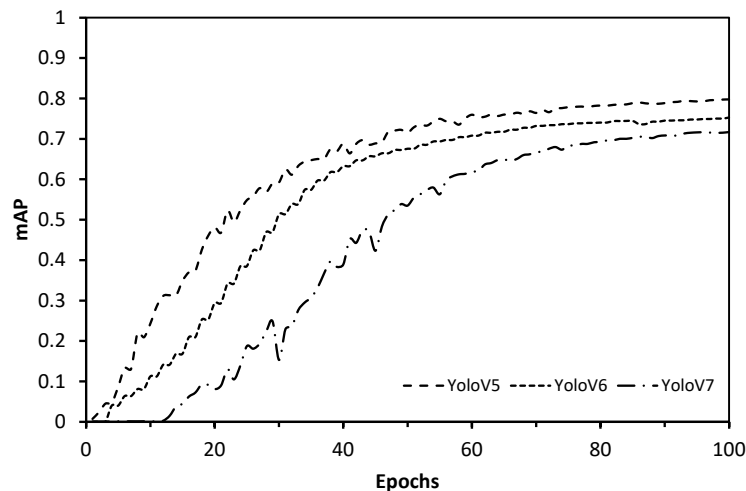


Figure 7. 7CGL mAP for YOLO models

As depicted in the Table 5 and Figure 8(a), the recall values for first three classes remains high. The lowest recall in this group is 0.98 which is produced by YOLOv6. It is worth mentioning that YOLOv6's recall value remains highly competitive with the best performing model, i.e., YOLOv5. However, as the object become smaller and away from the center of the image, YOLOv6 struggle to produce a competitive result. For example, recalls for 2nd molar are 0.95 and 0.84 for YOLOv5 and v6 respectively.

For all seven classes, YOLOv5 produced the overall highest F1 score as depicted in Figure 8(b). For the 1st and 2nd molar classes, the model significantly outperformed YOLOv6 and YOLOv7. However, the F1 score for these classes still remains comparatively low to other tooth classes. For instance, F1-score for incisor class is 0.95, however for 2nd Molar it drops to 0.78.

Table 5. 7CGL mAP, recall, and F1 scores for YOLO models

7CGL classes	mAP			Recall			F1-score		
	YOLOv5	YOLOv6	YOLOv7	YOLOv5	YOLOv6	YOLOv7	YOLOv5	YOLOv6	YOLOv7
Central Incisor	0.905	0.883	0.871	0.999	0.990	0.997	0.950	0.933	0.930
Lateral Incisor	0.885	0.866	0.851	0.999	0.990	0.989	0.939	0.924	0.915
Canine	0.869	0.854	0.828	0.997	0.990	0.984	0.929	0.917	0.899
1 st Premolar	0.808	0.784	0.745	0.992	0.980	0.987	0.891	0.871	0.849
2 nd Premolar	0.742	0.705	0.658	0.981	0.960	0.953	0.845	0.813	0.778
1 st Molar	0.718	0.647	0.584	0.953	0.920	0.842	0.819	0.760	0.690
2 nd Molar	0.663	0.530	0.482	0.952	0.840	0.708	0.782	0.650	0.574

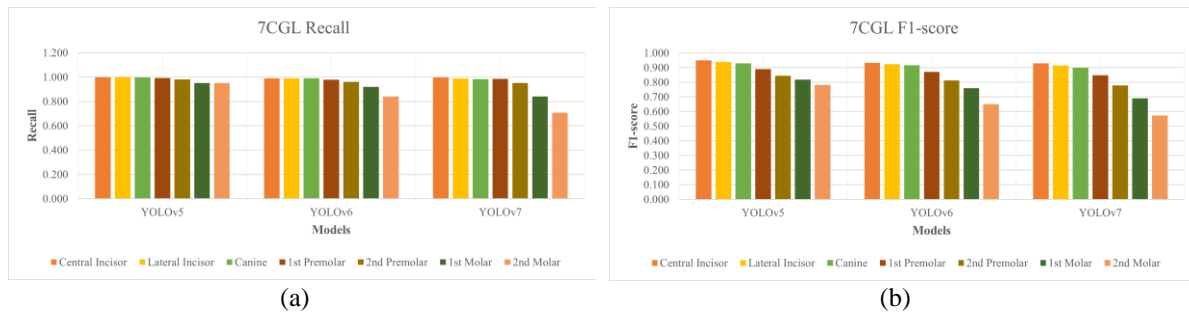


Figure 8. Results summary for 7CGL using YOLO models including (a) 7CGL recall and (b) 7CGL F1-score

For the first three classes, all models produced an F1 score that was equal to or greater than 0.90. These results indicate the strength of the YOLO architecture in detecting objects with high pixel representation and illumination levels. However, the F1 scores of models suffer greatly as the aforementioned features become less prominent. Which shows that YOLO architecture has limitations when it comes to detect small object with low illumination condition. Overall, it was found that YOLOv5 outperformed all other models for detecting teeth at the finest granular level (i.e., at the individual tooth level). According to the findings for 4CGL and 7CGL shown in Tables 4 and 5, YOLOv5 performed better than all other models in the benchmarking exercise and was chosen as the best model for the seven-class granular level classification challenge. In the context of 7CGL, YOLOv5 demonstrated its superiority over other state-of-the-art faster R-CNN models, achieving a mAP of 0.80, whereas faster R-CNN-101 yielded a significantly lower mAP of 0.54 at this level [10].

5. CONCLUSION AND FUTURE WORK

Utilizing a deep learning-based approach for tooth detection in intra-oral dental images can expedite the early diagnosis of tooth diseases and assist dental practitioners in identifying precise treatment options, ultimately leading to time and effort savings. Artifacts such as noises, occlusion, and overlapping pose significant challenges that can impede the accuracy of teeth detection and classification models. The presented study, evaluated the effect of systematic reduction in granularity on the performance improvement of deep learning models. Three models of YOLO named YOLOv5, YOLOv6, and YOLOv7 were employed for teeth detection and classification at three intra-oral teeth granularity levels. At level one, the performance of YOLOv6 was exceptional, demonstrating precise object localization as well as notable precision and recall. YOLO's strength is in detecting large objects, and its accuracy may suffer when it comes to capturing finer details within those objects. The difficulties arise when accurately detecting occluded or overlapping objects, because occlusion and overlap can obscure critical visual cues. Furthermore, handling objects of varying sizes and shapes can be difficult because it necessitates a more robust detection mechanism to accurately identify and delineate objects. These issues affected the performance of YOLO models as the granularity level was decreased. Despite the fact, the YOLOv5 still manage to accurately identify tooth objects at 0.89 and 0.79 mAP for level two and three respectively. In future, a hierarchical image representation which allows for capturing contextual information related to tooth structure, neighboring teeth, and the overall arrangement of the dental arch may be implemented. This approach may improve both the accuracy of detection and classification tasks.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude and appreciation to Universiti Teknologi PETRONAS for their generous provision of resources and materials, which were instrumental in the successful completion of this research work. This research was funded by the University Research Internal Fund (URIF), Project Cost Centre No. 015LB0-086.




REFERENCES

- [1] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," *Neural Networks*, vol. 32, pp. 333–338, 2012, doi: 10.1016/j.neunet.2012.02.023.
- [2] T. Sangpetch, T. Boongoen, and N. I. -On, "Profiling astronomical objects using unsupervised learning approach," *Computers, Materials and Continua*, vol. 74, no. 1, pp. 1641–1655, 2023, doi: 10.32604/cmc.2023.026739.




- [3] M. A. Amani and S. A. Sarkodie, "Mitigating spread of contamination in meat supply chain management using deep learning," *Scientific Reports*, vol. 12, no. 1, pp. 1-10, 2022, doi: 10.1038/s41598-022-08993-5.
- [4] S. AbuSalim, N. Zakaria, M. R. Islam, G. Kumar, N. Mokhtar, and S. J. Abdulkadir, "Analysis of deep learning techniques for dental informatics: a systematic literature review," *Healthcare*, vol. 10, no. 10, pp. 1-30, 2022, doi: 10.3390/healthcare10101892.
- [5] E. Kaya, H. G. Gunec, K. C. Aydin, E. S. Urkmez, R. Duranay, and H. F. Ates, "A deep learning approach to permanent tooth germ detection on pediatric panoramic radiographs," *Imaging Science in Dentistry*, vol. 52, no. 3, pp. 275-281, 2022, doi: 10.5624/isd.20220050.
- [6] S. K. Bayrakdar *et al.*, "A deep learning approach for dental implant planning in cone-beam computed tomography images," *BMC Medical Imaging*, vol. 21, no. 1, pp. 1-9, 2021, doi: 10.1186/s12880-021-00618-z.
- [7] F. P. Mahdi, K. Motoki, and S. Kobashi, "Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs," *Scientific Reports*, vol. 10, no. 1, pp. 1-12, Nov. 2020, doi: 10.1038/s41598-020-75887-9.
- [8] E. Kaya, H. G. Gunec, S. S. Gokyay, S. Kutal, S. Gulum, and H. F. Ates, "Proposing a CNN method for primary and permanent tooth detection and enumeration on pediatric dental radiographs," *Journal of Clinical Pediatric Dentistry*, vol. 46, no. 4, pp. 293-298, 2022, doi: 10.22514/1053-4625-46.4.6.
- [9] A. B. Oktay, "Tooth detection with Convolutional Neural Networks," in *2017 Medical Technologies National Conference, TIPTEKNO 2017*, IEEE, 2017, pp. 1-4, doi: 10.1109/TIPTEKNO.2017.8238075.
- [10] S. AbuSalim, N. Zakaria, S. A. Mostafa, Y. K. Hooi, N. Mokhtar, and S. J. Abdulkadir, "Multi-granularity tooth analysis via faster region-convolutional neural networks for effective tooth detection and classification," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 6, pp. 731-741, 2023, doi: 10.14569/ijacsa.2023.0140678.
- [11] D. Verma, S. Puri, S. Prabhu, and K. Smriti, "Anomaly detection in panoramic dental x-rays using a hybrid deep learning and machine learning approach," in *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, IEEE, Nov. 2020, pp. 263-268, doi: 10.1109/TENCON50793.2020.9293765.
- [12] G. H. Kwak *et al.*, "Automatic mandibular canal detection using a deep convolutional neural network," *Scientific Reports*, vol. 10, no. 1, pp. 1-8, 2020, doi: 10.1038/s41598-020-62586-8.
- [13] S. Abusalim, S. A. Mostafa, N. Zakaria, S. J. Abdulkadir, and N. Mokhtar, "Data augmentation on intra-oral images using image manipulation techniques," in *2022 International Conference on Digital Transformation and Intelligence, ICDI 2022 - Proceedings*, IEEE, 2022, pp. 117-120, doi: 10.1109/ICDI57181.2022.10007158.
- [14] H. Chen *et al.*, "A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films," *Scientific Reports*, vol. 9, no. 1, pp. 1-11, 2019, doi: 10.1038/s41598-019-40414-y.
- [15] D. V. Tuzoff *et al.*, "Tooth detection and numbering in panoramic radiographs using convolutional neural networks," *Dentomaxillofacial Radiology*, vol. 48, no. 4, pp. 1-10, 2019, doi: 10.1259/dmfr.20180051.
- [16] Y. Yasa *et al.*, "An artificial intelligence proposal to automatic teeth detection and numbering in dental bite-wing radiographs," *Acta Odontologica Scandinavica*, vol. 79, no. 4, pp. 275-281, Nov. 2021, doi: 10.1080/00016357.2020.1840624.
- [17] J. H. Lee, S. S. Han, Y. H. Kim, C. Lee, and I. Kim, "Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, vol. 129, no. 6, pp. 635-642, 2020, doi: 10.1016/j.oooo.2019.11.007.
- [18] S. Yilmaz, M. Tasyurek, M. Amuk, M. Celik, and E. M. Canger, "Developing deep learning methods for classification of teeth in dental panoramic radiography," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 2023, doi: 10.1016/j.oooo.2023.02.021.
- [19] M. Du, X. Wu, Y. Ye, S. Fang, H. Zhang, and M. Chen, "A combined approach for accurate and accelerated teeth detection on cone beam CT images," *Diagnostics*, vol. 12, no. 7, pp. 1-13, 2022, doi: 10.3390/diagnostics12071679.
- [20] M. D. N. Gerhardt *et al.*, "Automated detection and labelling of teeth and small edentulous regions on cone-beam computed tomography using convolutional neural networks," *Journal of Dentistry*, vol. 122, pp. 1-8, 2022, doi: 10.1016/j.jdent.2022.104139.
- [21] Y. Miki *et al.*, "Classification of teeth in cone-beam CT using deep convolutional neural network," *Computers in Biology and Medicine*, vol. 80, pp. 24-29, 2017, doi: 10.1016/j.combiomed.2016.11.003.
- [22] M. E. Celik, "Deep learning based detection tool for impacted mandibular third molar teeth," *Diagnostics*, vol. 12, no. 4, pp. 1-13, Apr. 2022, doi: 10.3390/diagnostics12040942.
- [23] C. Muramatsu *et al.*, "Tooth detection and classification on panoramic radiographs for automatic dental chart filing: improved classification by multi-sized input data," *Oral Radiology*, vol. 37, no. 1, pp. 13-19, 2021, doi: 10.1007/s11282-019-00418-w.
- [24] 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, doi: 10.1007/s11042-022-13644-y.
- [25] H. Lin and J. J. Yang, "Ensemble cross-stage partial attention network for image classification," *IET Image Processing*, vol. 16, no. 1, pp. 102-112, Sep. 2022, doi: 10.1049/ipr.2.12335.
- [26] A. Aljuaid and M. Anwar, "Survey of supervised learning for medical image processing," *SN Computer Science*, vol. 3, no. 4, pp. 1-22, 2022, doi: 10.1007/s42979-022-01166-1.
- [27] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2023, pp. 7464-7475, doi: 10.1109/cvpr52729.2023.00721.

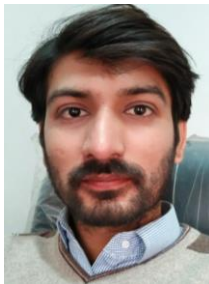
BIOGRAPHIES OF AUTHORS






Samah AbuSalim    a Ph.D. candidate in the Faculty of Computer Science & Information Technology, Universiti Teknologi PETRONAS (UTP), Malaysia. She is research assistant with the Department of Computer Science, Universiti Tun Hussein Onn Malaysia (UTHM), Malaysia. Before that, she obtained her Master of Computer Science degree at the Universiti Tun Hussein Onn Malaysia (UTHM), Malaysia in 2020. Her research interests include machine learning, computer vision, deep learning, software testing, and software quality. She can be contacted at email: samah_21000332@utp.edu.my.






Nordin Zakaria    is a senior lecturer at Department of Computer Information Sciences, Universiti Teknologi PETRONAS teaching. He obtained his Ph.D. from Universiti Sains Malaysia in 2007. His research interest includes computer graphics, agent-based simulation, and intelligent system. He can be contacted at email: nordinzakaria@utp.edu.my.






Aarish Maqsood    a researcher from Pakistan, is recognized for his extensive experience in machine learning, deep learning, geospatial analysis, and digital image processing. Over the years, expertise in Python, data visualization, and rainfall-runoff modeling has been honed by him. Significant contributions to climate change studies through publications have been made by him. He can be contacted at email: aarishmaqsood@gmail.com.






Abdul Saboor    received the B.S. degree in computer science from Bahria University, Islamabad, Pakistan, and the M.S. degree in Computer Science from COMSATS University Islamabad, Islamabad. He is currently pursuing the Ph.D. degree with the High-Performance Cloud Computing Centre, Universiti Teknologi PETRONAS, Malaysia. His research interests include cloud computing, big data analytics, and the internet of things. He can be abdul_19001745@utp.edu.my






Kwang Hooi Yew    is a senior lecturer at Department of Computer Information Sciences, Universiti Teknologi PETRONAS teaching and doing R&D in software engineering, ontology engineering and applications of AI. He had over 15 years of providing system development consultancy for PETRONAS Group Technology Services. Besides the routines, he authored several Oxford-Fajar Malaysia ICT textbooks at undergraduate, matriculations, polytechnic and secondary school levels in Malaysia and is a consulting author for Pelangi Publications and Praxis Publication Singapore. He can be contacted at email: yewkwanghooi@utp.edu.my.



Norehan Mokhtar    is a lecturer and consultant orthodontist at the Advanced Medical and Dental Institute, Universiti Sains Malaysia. She is also the Head of Dental Service of Universiti Sains Malaysia Bertam Medical Centre (PPUSMB) and the Head of Dental Simulation and Virtual Learning Research Excellence Consortium. Her research interest are cleft and craniofacial anomalies, orthodontics, biomaterials, and virtual learning. Her research findings were published in numerous citations indexed journals and her innovative research products won gold and silver medals at the research and innovation exhibition. She can be contacted at email: norehanmokhtar@usm.my.



Said Jadid Abdulkadir    (senior member, IEEE) received the B.Sc. degree in Computer Science from Moi University, the M.Sc. degree in computer science from Universiti Teknologi Malaysia, and the Ph.D. degree in information technology from Universiti Teknologi PETRONAS. He is currently an associate professor with the Department of Computer and Information Sciences, Universiti Teknologi PETRONAS. His current research interests include supervised machine learning and predictive and streaming analytics. He is currently serving as a journal reviewer for artificial intelligence review, IEEE ACCESS, and knowledge-based systems. He can be contacted at email: saidjadid.a@utp.edu.my.