**Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics** e-ISSN: 2656-8624
Homepage: https://ijeeemi.org/; Vol. 7, No. 2, pp. 253-269, April 2025

RESEARCH PAPER

OPEN ACCES

# The Role of U-Net Segmentation for Enhancing Deep Learning-based Dental Caries Classification

**Muhammad Keysha Al Yassar[1], Maya Fitria[1,2*], Maulisa Oktiana[1], Muhammad Aditya Yufnanda[1], Khairun Saddami[1], Kahlil Muchtar[1], and Teuku Reza Auliandra Isma[3]**

[1] Department of Electrical and Computer Engineering, Engineering Faculty, Universitas Syiah Kuala, Darussalam, Banda Aceh, Indonesia
[2] Interactive Intelligent Systems Research Group, Universitas Syiah Kuala, Darussalam, Banda Aceh, Indonesia
[3] Institute of Automation, University of Rostock, Rostock, Germany

## ABSTRACT

**Dental caries, one of the most prevalent oral diseases, can lead to severe complications if left untreated. Early detection is crucial for effective intervention, reducing treatment costs, and preventing further deterioration. Recent advancements in deep learning have enabled automated caries detection based on clinical images; however, most existing approaches rely on raw or minimally processed images, which may include irrelevant structures and noise, such as the tongue, lips, and gums, potentially affecting diagnostic accuracy. This research introduces a U-Net-based tooth segmentation model, which is applied to enhance the performance of dental caries classification using ResNet-50, InceptionV3, and ResNeXt-50 architectures. The methodology involves training the teeth segmentation model using transfer learning from backbone architectures ResNet-50, VGG19, and InceptionV3, and evaluating its performance using IoU and Dice Score. Subsequently, the classification model is trained separately with and without segmentation using the same hyperparameters for each model with transfer learning, and their performance is compared using a confusion matrix and confidence interval. Additionally, Grad-CAM visualization was performed to analyze the model's attention and decision-making process. Experimental results show a consistent performance improvement across all models with the application of segmentation. ResNeXt-50 achieved the highest accuracy on segmented data, reaching 79.17%, outperforming ResNet-50 and InceptionV3. Grad-CAM visualization further confirms that segmentation plays a crucial role in directing the model's focus to relevant tooth areas, improving classification accuracy and reliability by reducing background noise. These findings highlight the significance of incorporating tooth segmentation into deep learning models for caries detection, offering a more precise and reliable diagnostic tool. However, the confidence interval analysis indicates that despite consistent improvements across all metrics, the observed differences may not be statistically significant.**

## PAPER HISTORY

Received Feb. 02, 2025
Accepted April 20, 2025
Published April 23, 2025

## KEYWORDS

Dental caries;
Caries detection;
Teeth Segmentation;
U-Net;
Deep learning

## AUTHOR EMAIL

keysha@mhs.usk.ac.id
mayafitria@usk.ac.id

## 1. INTRODUCTION

Dental caries is a prevalent chronic infectious disease [1], defined by the demineralization of dental hard tissues such as enamel, dentine, and cementum in both primary and permanent teeth [2]. Caries can affect both children and adults [3]. When left untreated, caries can advance to form a cavitated lesion and may eventually affect the pulp, leading to pain, swelling, and potentially systemic symptoms [4]. Prolonged neglect can lead to significant destruction of the tooth's structural integrity and supporting tissues, resulting in irreversible tooth loss [5]. According to the latest Indonesian Basic Health Research in 2018, dental caries is among Indonesia's most widespread oral diseases, affecting 88.8% of individuals across all age groups. [6]. Therefore, early detection and precise caries diagnosis are essential to avoid serious damage and improve oral health outcomes.

Intraoral clinical images, obtained through photographic methods, are being recognized as a valuable resource for dental diagnostics, particularly in the context of automated image analysis using artificial intelligence [7][8]. Unlike radiographic imaging, which exposes patients to substantial levels of ionizing radiation [9], intraoral photographs are noninvasive, accessible, and cost-effective, making them ideal for population-wide caries screening and detection [10].

Various studies have been conducted on detecting dental caries using intraoral images. Fitria et al. developed an intraoral clinical image dataset for detecting dental caries and trained the model using ResNet50 with

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

various learning rates [8]. This study results in a dataset consisting of 700 images, augmented to increase the dataset size. The model achieved its highest accuracy of 79.52% with a learning rate of $10^{-4}$. J. K¨uhnisch et al. conducted a similar study utilizing an artificial intelligence-based Convolutional Neural Network (CNN) to classify clinical dental images into three categories: healthy teeth (caries-free), teeth with non-cavitated carious lesions, and teeth with cavitated caries [7]. Their approach employed the MobileNetV2 model, trained using transfer learning techniques. The results demonstrated that the CNN model based on MobileNetV2 exhibited high performance in detecting dental caries, achieving an overall accuracy of 92.5%. In 2024, Fitria et al. extended their research by developing a deep learning model for detecting decayed, missing, and filled teeth using YOLOv5 and YOLOv8 architectures. The YOLOv5l model, trained with a learning rate of $10^{-2}$, achieved a precision of 0.97, a recall of 0.858, and a mean average precision (mAP) of 0.904 in 1 hour and 18 minutes. Meanwhile, YOLOv8 demonstrated greater training stability, with larger variants performing better in detection tasks [11], [12].

Despite these advancements, existing approaches are based on raw or minimally processed clinical images, which can introduce irrelevant background information such as the tongue and lips. This extraneous information can hinder the model's ability to focus on specific features of the tooth, potentially affecting accuracy and reliability. Therefore, this study aims to develop and implement a U-Net-based teeth segmentation model to enhance dental caries detection by isolating relevant tooth structures and reducing the influence of background noise. This study contributes by developing a U-Net-based teeth segmentation model capable of isolating healthy teeth and visible caries lesions while masking out fillings, missing teeth, and irrelevant background information. This segmentation process enables further analysis in caries classification using deep learning. Additionally, a caries classification model is trained using three architectures—ResNet50, InceptionV3, and ResNeXt50—with and without segmentation to compare performance and analyze the impact of segmentation on model accuracy. To further understand the model's decision-making process, the Grad-CAM technique is used to visualize and compare classification results between original and segmented images, providing deeper insights into how segmentation influences classification outcomes.

This study is structured as follows: Section II provides an overview of the dataset, proposed methodology, and the training and evaluation methods used in this study. Section III presents the results of model training and evaluation for both segmentation and classification, including models trained with and without the segmentation process. Section IV offers a detailed discussion of the findings, covering result interpretation and the study's limitations. Finally, Section V concludes with a summary of the objectives, key findings, and potential directions for future research.

## 2. MATERIALS AND METHOD

Fig. 1 shows the processes undertaken in this study, starting with segmentation model development, followed by caries classification, and concluding with result analysis. The following subsection provides a detailed explanation of each process.

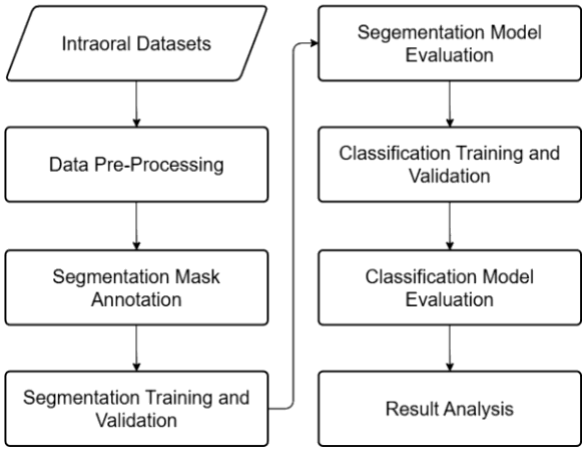



**Fig. 1. The research methodology flowchart illustrates the key stages of this study, from data preprocessing to result analysis.**

### A. Dataset

The datasets utilized in this study were adapted from those used previously by Fitria et al. in [8], with additional data collected from the Dental and Oral Polyclinic at Zainoel Abidin Regional Hospital in Banda Aceh. The dataset consists of dental clinical images of patients aged 18 to 45 years, captured using a Nikon D5300 (24MP) with an ISO range of 100–12,800 and an aperture of f/5.6. The images were taken under room lighting conditions, maintaining a fixed distance of 20 cm between the lens and the subject.

The dataset includes five intraoral sides, namely anterior (labial), right side (right buccal), left side (left buccal), inner upper teeth (upper occlusal), and inner lower teeth (lower occlusal). Examples of these views are illustrated in Fig. 2 (a), (b), (c), (d), and (e), respectively. The dataset consists of 1,274 images, including 721 normal teeth and 553 carious teeth, each with a resolution of 224x224 pixels.

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

**Fig. 2. Five Sides of the Intraoral Clinical Image Dataset, (a) Labial, (b) Right Buccal, (c) Left Buccal, (d) Upper Occlusal, (e) Lower Occlusal.**

### B. Hardware and Software Specifications

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

This study was conducted on a machine equipped with a GPU to accelerate the training process. The hardware specification and software environments are detailed in Table 1 to support the reproducibility of this study.

**Table 1. Hardware and Software Environment Used For Model Training and Development**

| Component | Specification |
|---|---|
| GPU | NVIDIA GeForce RTX |
| RAM | 24 GB DDR5 |
| Python | 3.10.14 |
| TensorFlow | 2.10.0 |
| cudatoolkit | 11.2.2 |
| cuDNN | 8.1.0.77 |

### C. Data Pre-Processing

To prepare the dataset for the training process, the images were first normalized using min-max normalization to scale pixel intensity values to a range of 0 to 1. This normalization step helps accelerate model convergence and ensures consistency in the input data [13]. After normalization, the dataset was divided into three groups: training, validation, and testing, with a ratio of 70:15:15 [14]. The split was performed randomly while maintaining the original class proportions in each subset to ensure balanced and unbiased data distribution, this splitting method is proven effective to handle an uneven dataset distribution [15]. A 180-degree rotation was applied to the training set as an augmentation technique. This process aimed to increase the variability of the training data and reduce the risk of overfitting [16]. Following the augmentation, the dataset's total size increased to 2165 images. Table 2 presents the final dataset distribution after the split and augmentation process.

**Table 2. The dataset distribution for training, validation, and testing shows the number of normal and caries cases in each set.**

| Datasets | Normal | Caries | Total |
|---|---|---|---|
| Training | 976 | 806 | 1782 |
| Validation | 117 | 74 | 191 |
| Testing | 116 | 76 | 192 |

### D. Segmentation Mask Annotation

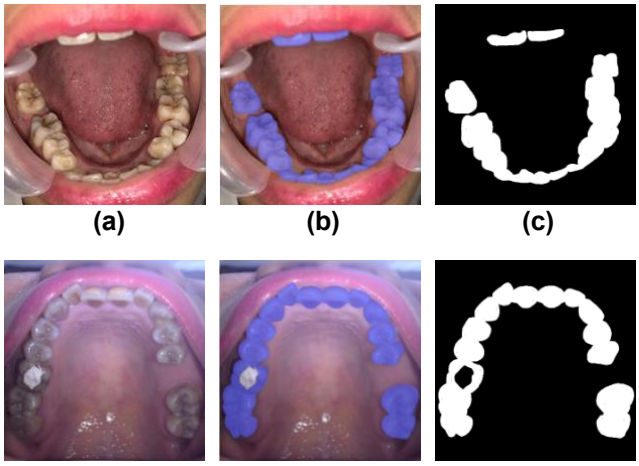
The mask annotation process was conducted manually using Label Studio to create the ground truth masks. Annotation was performed by drawing pixel-wise segmentation masks for the teeth in each image. Each intraoral clinical image was labeled to generate the segmentation masks for different conditions, namely, normal, decayed, filled, and missing teeth. These masks are used as ground truth for training and validating the segmentation model. The annotation process for normal and caries teeth is shown in Fig. 3 and Fig. 4, respectively, showcasing how regions of interest were delineated for each corresponding condition.

Specific approaches were taken for each condition:

- Normal: Masks were generated for intact and healthy tooth regions, as shown in Fig. 3, where the mask result highlights the complete tooth structures.
- Decay: Decayed areas were carefully segmented to highlight visible carious lesions, as demonstrated in Fig. 4.
- Filling: Regions with dental fillings were excluded, leaving only the actual tooth structures segmented (see Fig. 4(d), (e) and (f)).
- Missing: Edentulous areas without teeth were marked as absent to reflect missing regions accurately (see Fig. 4(d), (e) and (f)).

(a) (b) (c)

**Fig. 3. Visualization of normal teeth annotation, showing (a) original image, (b) annotation process, and (c) final annotation result.**



(a) (b) (c)

(d) (e) (f)

**Fig. 4. Visualization of caries teeth annotation, illustrating (a) and (d) the original images, (b) and (e) the annotation process, and (c) and (f) the final annotation results.**

### E. Segmentation Model Training and Validation

The U-Net architecture, widely used for medical image segmentation [17], [18], [19], is employed to train the teeth segmentation model. It consists of two main components: the contracting (encoder) and expanding (decoder) paths [20], [21]. The encoder captures spatial context through convolutional and pooling layers, which reduce the feature map size while increasing the number of channels [22]. The decoder restores spatial resolution using

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

transposed convolutions and skip connections, which link the encoder and decoder layers to retain fine details [23]. For feature extraction in the decoder, multiple backbones—ResNet50, VGG19, and InceptionV3—are utilized to leverage their advanced feature extraction capabilities using transfer learning from pre-trained ImageNet weights [24], [25]. The segmentation model is trained using Adam optimizer with a learning rate of $10^{-5}$ for 100 epochs, utilizing Intersection over Union Loss as the loss function, which has been proven to improve the model's ability on the segmentation task compared to regular CrossEntropy Loss [26]. The model's performance was evaluated during training using a validation process to ensure its effectiveness on unseen validation data. The complete architecture and parameters used for training the segmentation model are detailed in Table 3.

**Table 3. Hyperparameters used for segmentation model training to optimize model performance.**

| Parameter | Value |
|---|---|
| Backbones | ResNet50, VGG19, InceptionV3 |
| Loss Function | Intersection over Union Loss |
| Optimizer | Adam |
| Epoch | 100 |
| Learning Rates | $10^{-5}$ |
| Batch Size | 8 |

### F. Segmentation Model Evaluation

Evaluating semantic segmentation models involves both assessing classification accuracy and verifying how well the predicted regions align with the actual object boundaries. This requirement makes segmentation evaluation more complex than simple classification tasks [27]. To effectively capture both aspects of performance, Intersection over Union (IoU) and Dice Score are used as they quantify the overlap between the predicted segmentation and the ground truth mask [28]. The calculation of IoU is depicted in (Eq. (1)).

$$IoU = \frac{|S_{pred} \cap S_{gt}|}{|S_{pred} \cup S_{gt}|} \quad (1)$$

where $|S_{pred} \cap S_{gt}|$ is the number of overlapping pixels between the predicted and ground truth masks and $|S_{pred} \cap S_{gt}|$ is the total number of pixels in the union of predicted and ground truth masks. The IoU value ranges from 0 to 1, where a higher value indicates better performance in terms of the overlap between the predicted and ground truth masks. The calculation of the Dice Score is depicted in (Eq. (2)).

$$Dice\ Score = \frac{2|S_{pred} \cap S_{gt}|}{|S_{pred}| + |S_{gt}|} \quad (2)$$

where $|S_{pred} \cap S_{gt}|$ is the number of overlapping pixels between the predicted and ground truth masks, $|S_{pred}|$ is the total number of pixels in the predicted mask, and $|S_{gt}|$ describes the total number of pixels in the ground truth mask.

Similar to IoU, the Dice Score ranges from 0 to 1, where a higher value indicates a greater similarity between the predicted and ground truth masks. These metrics are particularly useful for evaluating medical image segmentation, as they emphasize overlapping regions [27], [29]. This evaluation process evaluates each segmentation model and identifies the best-performing model, which will be used for image segmentation in the classification stage.

### G. Classification Model Training and Validation

To develop the classification model, both the original and segmented images, where segmentation was performed using a previously trained teeth segmentation model, were trained separately using multiple model architectures, including ResNet50, InceptionV3, and ResNeXt50. These three architectures have been utilized to address dental image classification tasks, demonstrating strong performance due to their strong feature extraction capabilities [8], [30], [31]. Transfer learning was applied by initializing the models with pre-trained weights from ImageNet. All models were trained using the SGD optimizer with a learning rate of 2 x $10^{-5}$, using CrossEntropy Loss as the loss function, for 200 epochs to ensure a valid comparison. During training, a validation process was performed to assess the model's performance on the validation dataset. The optimal model weights were chosen based on the lowest validation loss to improve generalization. The complete architecture and parameters used for training the classification model are detailed in Table 4.

**Table 4. Hyperparameters used for classification model training, including backbone architectures, loss function, optimizer, number of epochs, learning rate, and batch size.**

| Parameter | Value |
|---|---|
| Architecture | ResNet50, InceptionV3, ResNeXt50 |
| Loss Function | CrossEntropy Loss |
| Optimizer | SGD |
| Epoch | 200 |
| Learning Rates | $2 \times 10^{-5}$ |
| Batch Size | 16 |

### H. Classification Model Evaluation

The performance of the caries classification model is evaluated using the confusion matrix, categorizing predictions into four distinct outcomes. True Positives (TP) represent cases where caries are correctly identified, while True Negatives (TN) indicate correctly classified healthy teeth. False Positives (FP) occur when healthy teeth are mistakenly labeled as caries, leading to a Type I error or false alarm. Conversely, False Negatives (FN) refer to instances where caries are incorrectly classified as normal, resulting in a Type II error or missed detection

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

[32]. The confusion matrix plays a crucial role in assessing both correct and incorrect predictions, ensuring classification reliability, and reducing the risk of misdiagnosis, which could lead to inappropriate treatment decisions.

Five performance metrics are computed by analyzing these components to provide a comprehensive evaluation of the model, namely, accuracy, precision, specificity, recall, and F-1 score [33]. Accuracy (Eq. (3)) represents the proportion of correctly classified cases relative to the total number of samples. Precision (Eq. (4)) represents the proportion of correctly identified caries cases among all instances predicted as caries. Specificity (Eq. (5)) measures the model's ability to correctly classify normal teeth by calculating the proportion of actual normal cases correctly identified. Recall (Eq. (6)) indicates the proportion of actual caries cases that the model correctly detects. F1-Score (Eq. (7)) is the harmonic mean of precision and recall, providing a balanced measure of a model's ability to detect caries while minimizing false positives and false negatives.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

To assess the statistical significance of segmentation on caries classification performance, this study employs confidence intervals as part of the evaluation process. The confidence intervals for classification metrics are computed using the bootstrap-t method. Specifically, the test set is resampled randomly with replacement 1,000 times, with each resample being the same size as the original test set. This method allows for accurate and less biased estimation of confidence intervals [34]. Confidence intervals are calculated for the difference in metrics between the segmented and original datasets, enabling a direct assessment of whether segmentation results in statistically meaningful improvements. The 95% confidence interval calculation is presented in (Eq. (8))

$$\left(\hat{\Delta} - \hat{t}^{(1-\alpha)} \times \widehat{se}, \hat{\Delta} - \hat{t}^{(\alpha)} \times \widehat{se}\right) \quad (8)$$

where $\hat{\Delta}$ denotes the observed difference in performance metric, and $\widehat{se}$ represents the estimated standard error of the metric. The terms $\hat{t}^{(\alpha)}$ and $\hat{t}^{(1-\alpha)}$ are the $\alpha$ and $1-\alpha$ quantiles of the distribution of bootstrapped t-values, respectively.

The Grad-CAM (Gradient-weighted Class Activation Mapping) technique is also applied to visualize the classification model prediction [35], [36]. By leveraging the gradients of a specific class flowing into the final convolutional layer, this method generates a localization map that highlights the key regions the model focuses on [37]. The Grad-CAM is generated for the highest predicted class using the last convolutional layer of each architecture. Specifically, "conv5_block3_out" is used for ResNet50 and ResNeXt50, while "mixed10" is used for InceptionV3. The gradient of the predicted class score is computed with respect to these feature maps, and the ReLU activation function is applied to highlight the features that positively influence the model's decision. In the visualization, the regions that contribute most to the model's prediction are highlighted more prominently in red. This visualization helps illustrate the impact of segmentation on the model's focus, providing deeper insights into how the segmentation process influences the classification model's decision-making process [38].

The proposed methodology could be extended by incorporating larger and more balanced dataset distribution to improve generalization. Additionally, exploring more advanced deep learning architecture may enhance performance further.

## 3. RESULTS

### A. Segmentation Training and Validation Results

The training progression of the ResNet50-UNet model, as illustrated in Fig. 5 (a), indicates a well-converging process. The training loss decreases rapidly within the first 20 epochs before stabilizing at 0.0244. A similar trend is observed in the validation loss, which follows a steady downward trajectory and settles at 0.0782. The Intersection over Union (IoU) metric also improves significantly during the initial training phase, with the training IoU reaching 0.984 by the final epoch. The validation IoU follows a comparable pattern, rising sharply early on before stabilizing at 0.952.

The learning curves for the VGG19-UNet model, shown in Fig. 5 (b), display a smooth training progression. The training loss declines sharply in the initial 20 epochs, then decreases to a final value of 0.0279. The validation loss follows a similar pattern, eventually stabilizing at 0.0693. The IoU metric exhibits a comparable trend, with training and validation IoU increasing rapidly at the beginning of training. By the final epoch, the training IoU reaches 0.982, while the validation IoU stabilizes at 0.956.

The training curves for the InceptionV3-UNet model, shown in Fig. 5 (c), follow a pattern similar to the previous models. The training loss drops quickly within the first 20 epochs before leveling at 0.0285. Likewise, the validation loss steadily decreases and eventually stabilizes at 0.0778. The IoU metric shows a rapid increase early in training, with the training IoU reaching 0.981 and the validation IoU settling at 0.952.

Table 5 summarizes the training and validation results for each trained segmentation model. As observed from the metrics presented, the performance across all models is relatively similar, with only marginal differences at the decimal level. These results highlight that all three models effectively learn the segmentation task with minimal risk of overfitting, as indicated by the close alignment of

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

training and validation metrics. Notably, the VGG19-UNet model performs best, exhibiting the lowest validation loss (0.0693) and the highest validation IoU (0.956). This indicates that the VGG19-UNet model not only excels in learning the segmentation task but also generalizes well to unseen data.

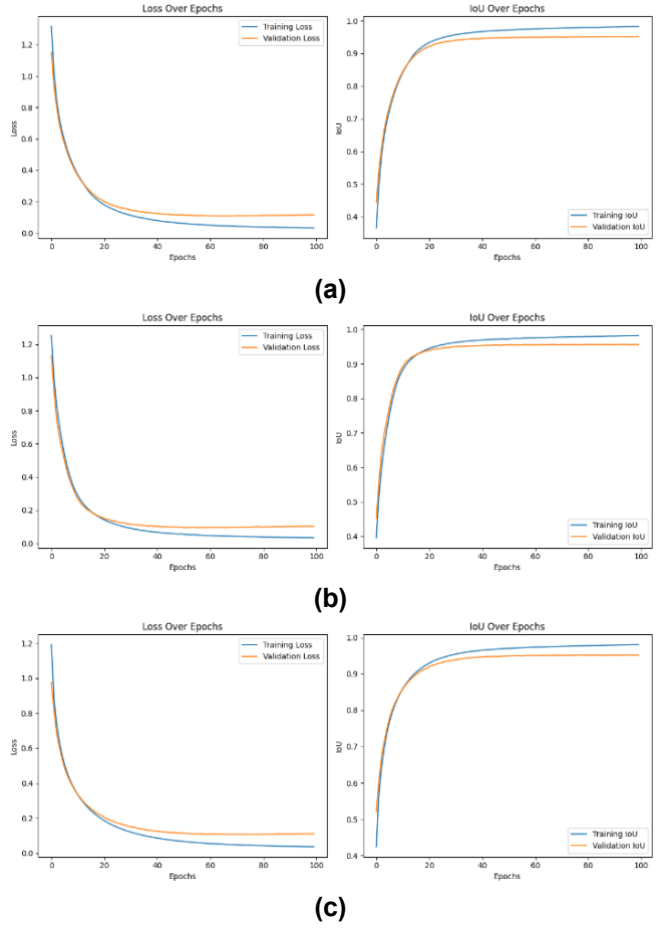



**Fig. 5. Training and validation curves of ResNet50-UNet (a), VGG19-UNet (b), and InceptionV3-UNet (c). The left plot shows the loss reduction over epochs, while the right plot illustrates the improvement in the Intersection over Union (IoU) metric.**

### B. Segmentation Model Evaluation Result

Table 6 exhibits the testing results of the three segmentation models, highlighting their Intersection over Union (IoU) and Dice Score performances. Overall, the results indicate that all three models achieve highly comparable performance, with only minor variations in their evaluation metrics. The InceptionV3-UNet model attains the highest test IoU at 0.9528, followed closely by ResNet50-UNet (0.9514) and VGG19-UNet (0.9508). Similarly, the test Dice Score metric exhibits marginal differences, with InceptionV3-UNet achieving the highest score of 0.9602, slightly outperforming VGG19-UNet (0.9600) and ResNet50-UNet (0.9587). Despite these minimal variations, the InceptionV3-UNet model is the most suitable choice based on its superior testing performance. Given its ability to achieve the highest IoU and Dice Score among the three models, InceptionV3-UNet is selected as the most optimal model and will be used for segmentation in the subsequent classification process.

**Table 5. Segmentation training and validation result (train loss, validation loss, train IoU and validation IoU) to compare performance among the three model.**

| Model | Train Loss | Val Loss | Train IoU | Val IoU |
|---|---|---|---|---|
| ResNet50 -UNet | 0.0244 | 0.0782 | 0.984 | 0.952 |
| VGG19-UNet | 0.0279 | 0.0693 | 0.982 | 0.956 |
| Inception V3-UNet | 0.0285 | 0.0778 | 0.981 | 0.952 |

**Table 6. Segmentation evaluation result (testing dice score and IoU) to compare performance among the three model.**

| Model | Test IoU | Test Dice Score |
|---|---|---|
| ResNet50-UNet | 0.9514 | 0.9587 |
| VGG19-UNet | 0.9508 | 0.9600 |
| InceptionV3-UNet | 0.9528 | 0.9602 |

Fig. 6 presents examples of segmentation predictions generated by the three trained models. Overall, all models accurately segment the dental regions for both normal and carious teeth, with the segmentation results being highly similar across models, exhibiting only minor variations. For normal teeth, each model effectively delineates the entire tooth structure with high precision, ensuring that all visible teeth are distinctly segmented from the background (see Fig. 6 (g)–(i), (m)–(o), (s)–(u)). In the case of carious teeth, the models successfully differentiate between healthy tooth structures and areas affected by dental restorations or missing teeth (see Fig. 6 (j) and (k), (p) and (q), (v) and (w)) while still including teeth with carious lesions, ensuring that the lesions remain visible in the segmented outputs (see Fig. 6 (i), (r), and (x)). This consistency across different models suggests their robustness in handling variations in dental conditions while preserving the essential features required for the classification process.

### C. Classification Model Training and Evaluation Result

Fig. 7 (a) presents the learning curves for ResNet-50 trained on the original dataset. The training loss steadily declines to 0.3917, while validation loss drops sharply in the first 50 epochs before stabilizing at 0.6241. Training accuracy consistently improves, reaching 84.79%, whereas validation accuracy rises rapidly early and settles at 70.68%. The best model weights were obtained at epoch 149, with the lowest validation loss of 0.6207 and validation accuracy of 69.63%. In contrast, Fig. 7 (b)

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

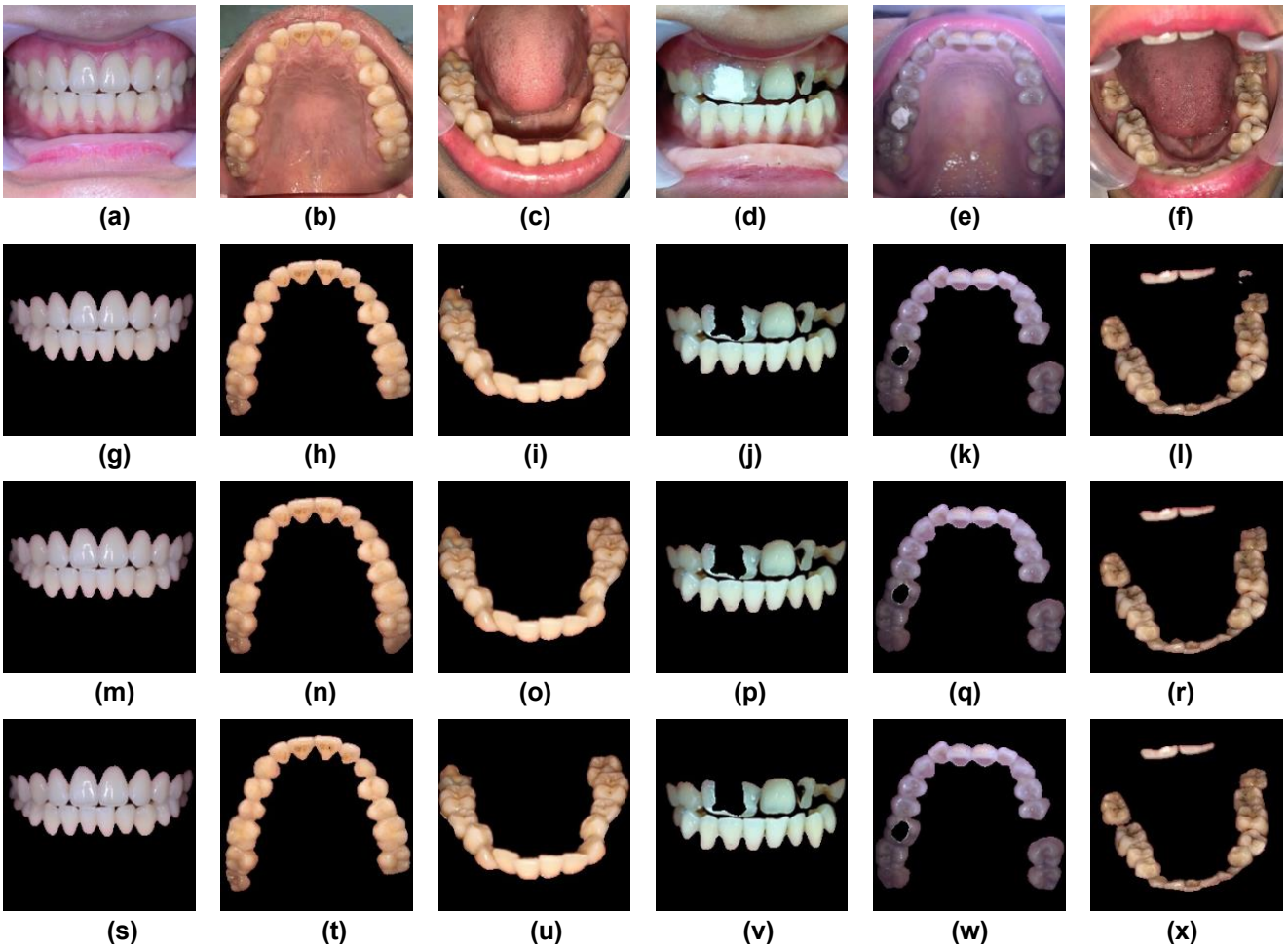



**Fig. 6. Segmentation Model Prediction Result, (a-c) Original Normal Image, (d-f) Original Caries Image, (g-i) ResNet50-UNet Normal Image Segmentation, (j-l) ResNet50-UNet Caries Image Segmentation, (m-o) VGG19-UNet Normal Image Segmentation, (p-r) VGG19-UNet Caries Image Segmentation, (s-u) InceptionV3-UNet Normal Image Segmentation, (v-x) InceptionV3-UNet Caries Image Segmentation**

shows that training on the segmented dataset results in a more stable improvement. The training loss decreases to 0.3276, and validation loss drops sharply in the first 30 epochs before stabilizing at 0.5206. Training accuracy peaks at 89.17%, while validation accuracy jumps early and stabilizes at 72.77%. Notably, the model performs best at the final epoch, where validation loss is lowest.

Fig. 7 (c) illustrates the learning curves for InceptionV3 trained on the original dataset. The training loss steadily decreases to 0.4482, while validation loss drops sharply in the first 20 epochs before stabilizing at 0.6763. Training accuracy improves rapidly within the first 30 epochs and gradually reaches 81.20%. Validation accuracy follows a similar pattern, rising significantly early on and settling at 67.02%. The best model weights were obtained at epoch 104, where validation loss reached its lowest at 0.6721, with a validation accuracy of 66.49%.

As shown in Fig. 7 (d), training on the segmented dataset results in a more consistent improvement. The training loss decreases to 0.4014, while validation loss drops sharply in the first 30 epochs before stabilizing at 0.5609. Training accuracy rises quickly in the early stages and peaks at 83.78%. Similarly, validation accuracy improves significantly in the first 30 epochs before leveling off at 70.16%. The best model performance is achieved at epoch 181, with the lowest validation loss of 0.5587 and validation accuracy of 71.20%.

Fig. 7 (e) illustrates the learning curves for ResNeXt-50 trained on the original dataset. The training loss steadily declines, reaching 0.3463 by the final epoch, while validation loss drops within the first 25 epochs before stabilizing at 0.5880. Training accuracy consistently improves, peaking at 87.15%. Validation accuracy follows a steady upward trend until epoch 100, after which it plateaus at 71.73%. The model's best performance occurs at epoch 198, where validation loss reaches its lowest point at 0.5868, with a validation accuracy of 71.73%.

As depicted in Fig. 7 (f), training on the segmented dataset enhances model performance. The training loss decreases further to 0.2664, while validation loss shows a sharp decline in the early epochs before gradually

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

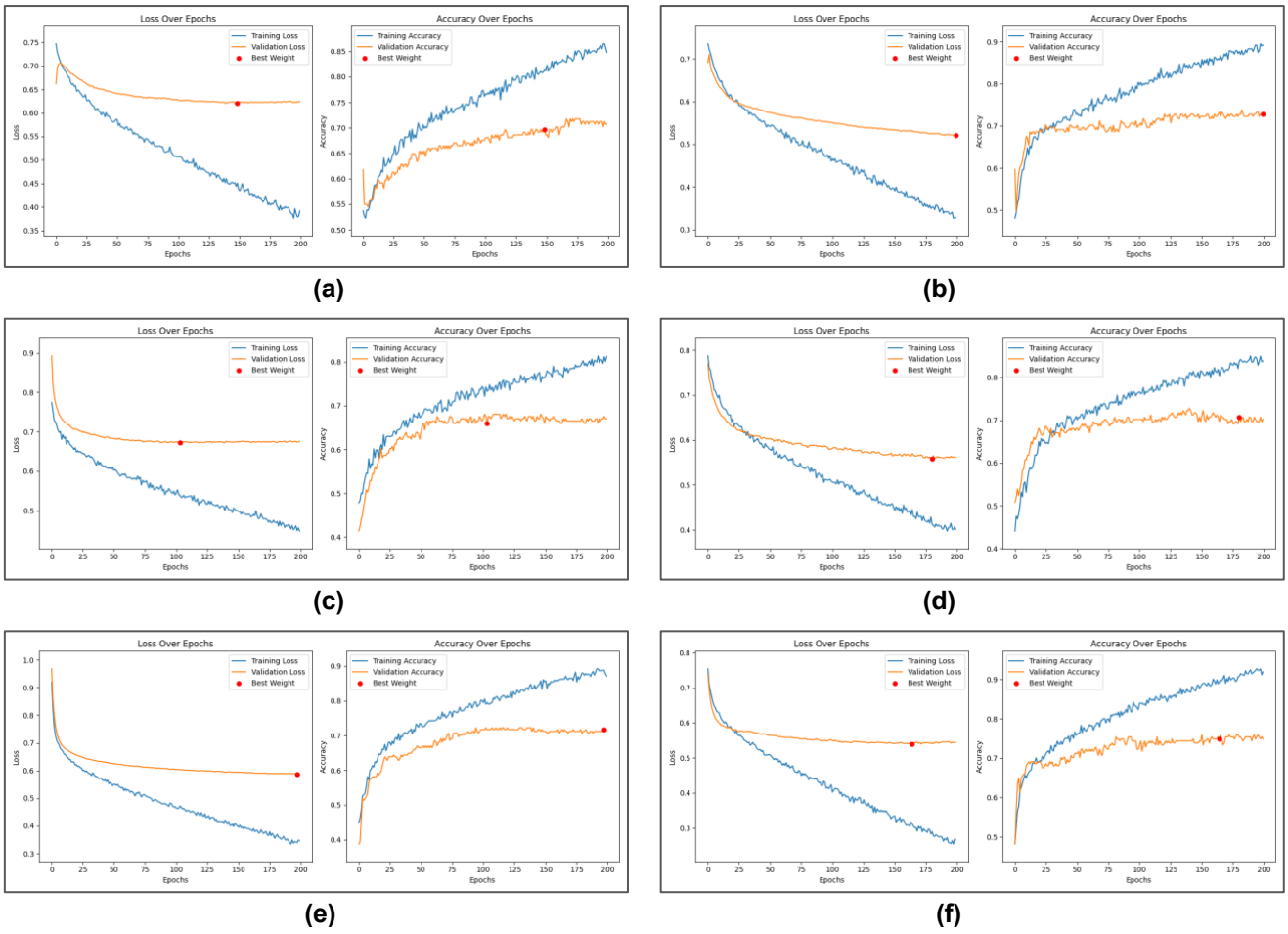



**Fig. 7. Classification Model Learning Curve, (a) ResNet50 Original Images, (b) ResNet50 Segmented Images, (c) InceptionV3 Original Images, (d) InceptionV3 Segmented Images, (e) ResNeXt50 Original Images, (f) ResNeXt50 Segmented Images**

stabilizing at 0.5446. Training accuracy rises quickly, culminating at 91.92%, while validation accuracy increases significantly in the initial stages and eventually plateaus at 74.87%. The model achieves optimal performance at epoch 165, where validation loss is minimized at 0.5393, with a validation accuracy of 75.39%.

Table 7 summarizes the classification performance of different architectures trained on original and segmented datasets. Acr oss all models, training on the segmented dataset consistently led to better training and validation performance compared to the original dataset, indicating that the models were able to learn more effectively from the segmented data. The best overall performance was achieved by the ResNeXt50 model trained on the segmented dataset, which recorded the lowest training loss (0.2252), the highest training accuracy (95.45%), and the highest validation accuracy (75.39%). Meanwhile, the lowest validation loss was observed in the ResNet50 model trained on the segmented dataset, reaching 0.253. However, despite having the lowest validation loss, its validation accuracy (72.77%) was still lower than that of ResNeXt50, suggesting that ResNeXt50 had a better balance between loss minimization and generalization.

**Table 7. Classification training and validation results (train loss, validation loss, train accuracy, and validation accuracy) to compare performance among the three models trained using original and segmented datasets.**

| Model | Dataset | Train Loss | Val Loss | Train Acc | Val Acc |
|---|---|---|---|---|---|
| ResNet50 | Original | 0.3902 | 0.6207 | 85.69 % | 69.63 % |
| | Segmented | 0.253 | 0.5206 | 94.39 % | 72.77 % |
| InceptionV3 | Original | 0.4876 | 0.6721 | 78.96 % | 66.49 % |
| | Segmented | 0.3284 | 0.5587 | 90.01 % | 71.20 % |
| ResNeXt50 | Original | 0.2659 | 0.5868 | 94.73 % | 71.73 % |
| | Segmented | 0.2252 | 0.5393 | 95.45 % | 75.39 % |

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

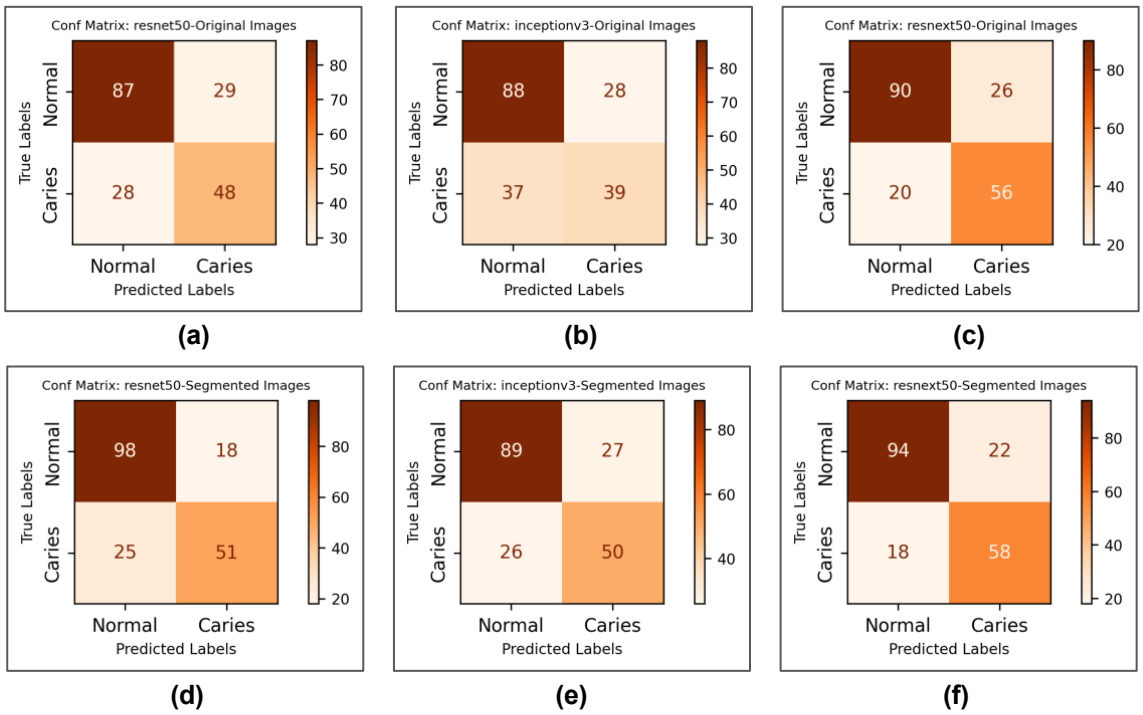



**Fig. 8. Classification Model Confusion Matrix, (a) ResNet50 Original Images, (b) InceptionV3 Original Images, (c) ResNeXt50 Original Images, (d) ResNet50 Segmented Images, (e) InceptionV3 Segmented Images, (f) ResNeXt50 Segmented Images**

### D. Classification Model Evalutaion Result

Fig. 8 presents the confusion matrices for the classification models trained on both original and segmented images, highlighting the impact of data preprocessing on model performance. Across all architectures, training with segmented images led to improved classification results. For ResNet50, the number of correctly classified normal cases increased from 87 to 98, while misclassified normal cases decreased from 29 to 18 (see Fig. 8 (a), and (d)). Similarly, the correctly classified caries cases slightly improved from 48 to 51, with false negatives decreasing from 28 to 25. InceptionV3 exhibited a similar trend, with correctly classified normal cases increasing from 88 to 89, while the classification of caries cases improved from 39 to 50 (see Fig. 8 (b) and (e)). Notably, ResNeXt50 demonstrated the most significant improvement when trained with segmented images, with correctly identified normal cases rising from 90 to 94 and caries cases increasing from 56 to 58. Additionally, false negatives decreased from 20 to 18, indicating better sensitivity to caries detection (see Fig. 8 (c) and (f)).

The evaluation metrics, including accuracy, precision, specificity, recall, and F1-score, were derived from the confusion matrix for each model. Table 8 presents the calculated values for each metric, offering a detailed comparison of the models trained on both original and segmented images. Across all architectures, models trained on segmented images consistently performed better than those trained on original images. Improvements in accuracy, precision, specificity, recall, and F1-score were observed, indicating that segmentation enhances feature extraction and reduces irrelevant background noise. Among the models, ResNeXt50 trained on segmented images achieved the highest accuracy (79.17%) and F1-score (79.25%), making it the best-performing model. The most significant improvement is observed in ResNet50, where accuracy increased from 70.31% to 77.60% after using segmented images. InceptionV3 showed the lowest overall performance among the three models, with an accuracy of 66.15% on original images, which improved to 72.40% after segmentation.

Fig. 9 presents the 95% confidence interval differences between the performance metrics of models trained on segmented versus original images. While the mean differences suggest an overall positive shift across most metrics, all intervals include zero value. These results indicate that although segmentation tends to yield positive trends across metrics, they may not consistently reflect statistically robust improvements.

Fig. 10 compares the Grad-CAM visualization of misclassified cases in the original dataset that were corrected after applying segmentation. The model trained on the original dataset focused on non-dental areas like lips, gums, and the palate, leading to incorrect classifications. With segmentation, its attention shifted to the teeth area, improving classification accuracy. This highlights the role of segmentation in removing irrelevant background information and enhancing caries

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

**Table 8. Classification evaluation results (accuracy, precision, specificity, recall, and F1-score) to compare performance among the three classification models trained using original and segmented datasets.**

| Model | Dataset | Accuracy | Precision | Specificity | Recall | F1-Score |
|---|---|---|---|---|---|---|
| ResNet50 | Original | 70.31% | 70.38% | 75.00% | 70.31% | 70.35% |
| | Segmented | 77.60% | 77.39% | 84.48% | 77.60% | 77.39% |
| InceptionV3 | Original | 66.15% | 65.57% | 75.86% | 66.15% | 65.71% |
| | Segmented | 72.40% | 72.46% | 76.72% | 72.40% | 72.43% |
| ResNeXt50 | Original | 76.04% | 76.46% | 77.59% | 76.04% | 76.18% |
| | Segmented | 79.17% | 79.40% | 81.03% | 79.17% | 79.25% |

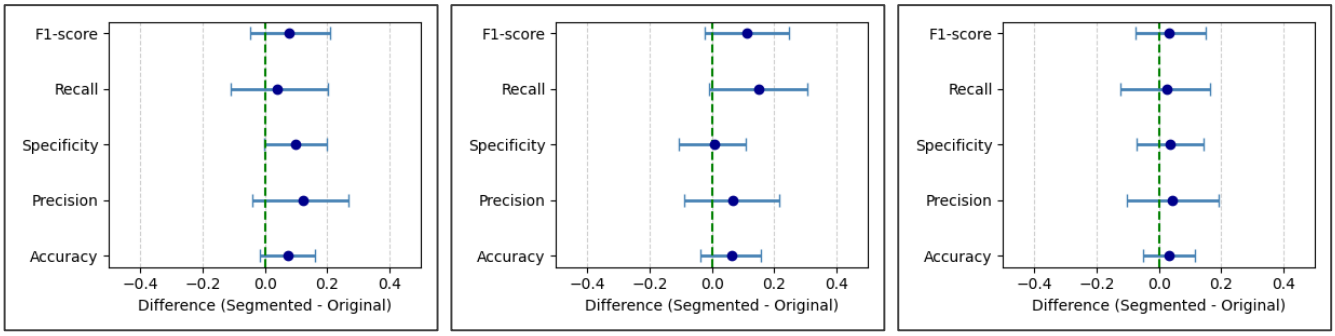



**Fig. 9. Difference in 95% confidence intervals of accuracy, precision, specificity, recall, and F1-score between models trained on segmented and original images across three architectures: (a) ResNet50, (b) InceptionV3, (c) ResNeXt50.**

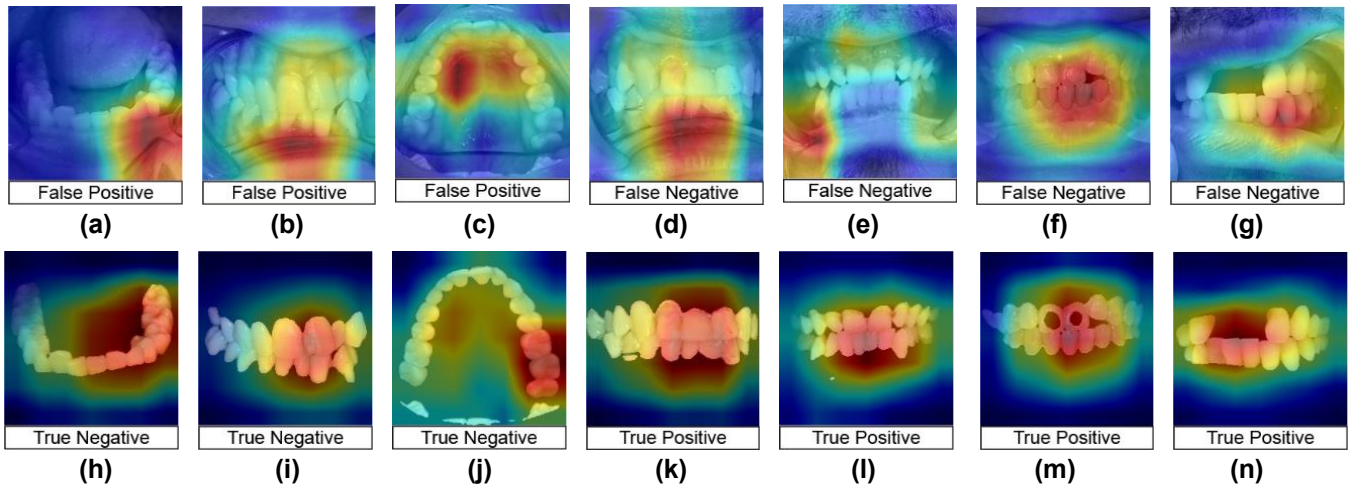



**Fig. 10. Grad-CAM visualization comparing misclassified cases in the original dataset (a-g), with corrected classifications in the segmented dataset (h-n)**

classification. Fig. 11 presents the Grad-CAM visualization of correctly classified cases from both datasets. While the model correctly identified caries in the original dataset, its focus included non-dental areas, which compromised the validity of its predictions. In contrast, segmentation helped the model concentrate on the relevant features, improving the model's reliability.

## 4. DISCUSSION

This study aims to develop a U-Net-based dental segmentation model to improve the accuracy of dental caries classification. The application of teeth segmentation techniques aims to allow the model to focus on the relevant tooth regions, reducing background noise and enhancing feature extraction for more accurate dental caries classification.

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

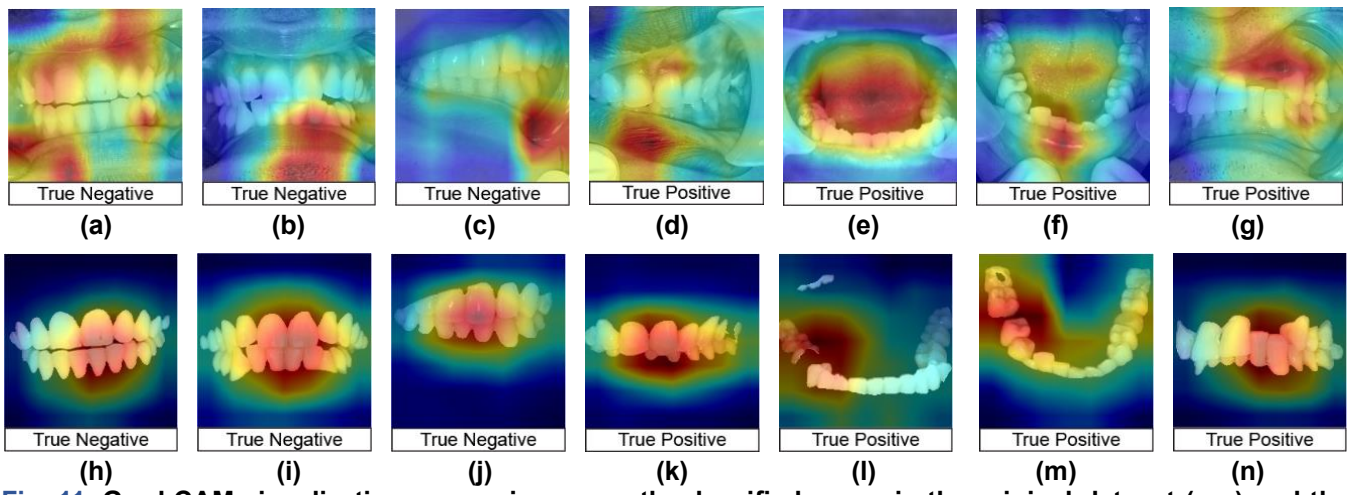



**Fig. 11. Grad-CAM visualization comparing correctly classified cases in the original dataset (a-g) and the segmented dataset (h-n). Both rows represent images that were correctly classified, highlighting differences in model focus between the original and segmented datasets.**

The results of the segmentation model development showed that combining U-Net with transfer learning using different backbones (ResNet50, VGG19, and InceptionV3) achieved high performance in segmenting teeth in the Intraoral Clinical Images dataset. As shown in Fig. 5 (a), (b), and (c), the training process demonstrated that U-Net effectively learned segmentation areas, with consistent improvement in both training and validation data. The minimal gap between training and validation performance indicates that the model learned well from the training data and generalized effectively to validation data without signs of overfitting. The evaluation results on the test dataset, as presented in Table 5, show that all three trained models achieved IoU scores of around 0.95 and Dice scores of around 0.96. These high values indicate that the segmentation models successfully identified the tooth regions with high overlap with the ground truth masks, ensuring that important dental structures were accurately segmented. The segmentation outputs in Fig. 6 demonstrate that all three models effectively distinguished healthy teeth, edentulous areas (missing teeth), and dental fillings. Additionally, the models were able to include and highlight caries lesions, ensuring that caries-affected areas remained visible for further classification.

Among all the models, InceptionV3-UNet produced the most promising results, achieving the highest IoU and Dice scores. This suggests that InceptionV3's feature extraction capabilities were particularly effective for this segmentation task, making it the best-performing segmentation backbone in this study. The following section discusses how integrating segmentation into the classification pipeline influenced the overall model accuracy and effectiveness in detecting caries.

Fig. 7 compares the training curves of models trained on the original and segmented datasets for each architecture. Fig. 7 (a) and Fig. 7 (b) specifically show the training curves for ResNet50 when trained on both datasets. In the training curve for the original dataset, the lowest validation loss was reached at epoch 149. After this point, the validation loss became stable but slowly increased, while the training loss continued to decrease steadily until the final epoch. This indicates that after epoch 149, the model began memorizing the training data rather than learning meaningful patterns. On the other hand, when trained on the segmented dataset, the lowest validation loss was achieved at the final epoch. This indicates that the model continued to learn useful features throughout the training process instead of simply memorizing the training data.

A similar pattern was observed with the InceptionV3 model (see Fig. 7 (c) and Fig. 7 (d)). When trained on the original dataset, the lowest validation loss was reached at epoch 104, gradually increasing while the training loss continued to decrease. In contrast, when trained on the segmented dataset, the lowest validation loss was achieved at epoch 181 and remained stable afterward. This suggests that the InceptionV3 model trained on the segmented dataset could continue learning important features from the training data more effectively until the final epoch compared to the model trained on the original dataset.

A different trend was observed in the ResNeXt50 model (see Fig. 7 (e) and Fig. 7 (f)). When trained on the original dataset, the validation loss decreased steadily until the final epoch. However, when trained on the segmented dataset, the lowest validation loss was reached at epoch 165, after which it gradually stabilized and increased slightly while the training loss continued to decrease. This indicates that, in the case of ResNeXt50, the model trained on the segmented dataset started memorizing the training data after epoch 165 and no longer learned useful features. Despite this, the ResNeXt50 model trained on the segmented dataset

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

achieved a relatively low validation loss (0.5393), only slightly higher than ResNet50 trained on the segmented dataset (0.5206), and obtained the highest validation accuracy among the three models at 75.39%.

The classification model evaluation performance in Table 8 demonstrates that training on the segmented dataset consistently improves performance across all three models compared to training on the original dataset. This improvement is reflected in higher accuracy, precision, recall, specificity, and F1-score, indicating that segmentation helps the models better distinguish between caries and normal teeth. Among the three models, ResNet50 shows a substantial increase in accuracy from 70.31% to 77.60%, along with improvements in precision (70.38% to 77.39%) and recall (70.31% to 77.60%). The increase in specificity from 75.00% to 84.48% suggests that the segmented dataset allows the model to reduce false positives. This significant gain may be attributed to ResNet50's simpler residual architecture, which benefits more from segmentation by suppressing background noise and enhancing its feature learning on caries-related areas. InceptionV3, which initially had the lowest accuracy at 66.15%, shows only moderate improvements with segmentation, reaching 72.40% accuracy. While recall also increases (66.15% to 72.40%), the specificity shows minimal gain (75.86% to 76.72%), indicating that the model still struggles to avoid false positives. This limited improvement may stem from InceptionV3's more complex multi-branch architecture, which spreads attention across multiple scales and may be less suited to the localized nature of caries features in full intraoral images. ResNeXt50 achieves the best overall performance, reaching the highest accuracy of 79.17% and recall of 79.17% after segmentation. Precision and specificity also improve to 79.40% and 81.03%, respectively, suggesting strong discriminative capability. However, the improvement margin is smaller compared to ResNet50, indicating that ResNeXt50's grouped convolutional structure is already effective at extracting robust features even from unsegmented images.

Fig. 10 compares Grad-CAM visualizations between misclassified cases on the original dataset and their corrected classifications on the segmented dataset. This visualization provides deeper insight into how segmentation influences the model's prediction process. In the visualized results, the intensely red regions indicate the areas the model focuses on the most, contributing significantly to its decision-making. In Fig. 10 (a) and 10 (h), as well as Fig. 10 (b) and 10 (i), the model's attention in the original images is primarily directed toward the lips, leading to the misclassification of normal teeth as carious. This indicates that, in the original images, the model fails to learn relevant caries-related features, relying instead on irrelevant background information. However, when using segmented images, the model's focus shifts correctly to the healthy tooth region, allowing it to classify the normal teeth accurately. In Fig. 10 (c) and 10 (j), which also depict normal teeth, the model in the original image heavily relies on the palate for classification, leading to misclassification. In contrast, in the segmented images, the most significant contribution to the model's prediction comes from the healthy teeth, specifically the upper posterior teeth, allowing the model to classify them as normal correctly. Fig. 10 (d) and 10 (k), along with Fig. 10 (e) and 10 (l), show teeth with carious lesions. In the original images, the model fails to focus on the tooth area, leading to the misclassification of carious teeth as normal. In contrast, with the segmented dataset, the model correctly directs its attention to the affected tooth regions, particularly the carious lesions, resulting in accurate classification. In Fig. 10 (f) and 10 (m), which feature a filled tooth, the model in the original image correctly focuses on the tooth structure but still misclassifies it as a normal tooth. Potentially due to the visual characteristics of a filled tooth closely resembling those of a normal tooth. However, using the segmented dataset, which excludes the filling area, helps the model to distinguish the dental filling from healthy teeth, allowing for the correct classification as carious. In the case of missing teeth (see Fig. 10 (g) and 10 (n)), the model in the original image is distracted by the intact tooth structure, leading to misclassification as a normal tooth. However, the model successfully focuses on the missing tooth area when using segmented images, allowing for correct classification. These comparison results demonstrate that the segmentation method helps the model to focus on relevant tooth structures, thereby improving classification accuracy.

Fig. 11 presents the Grad-CAM visualization of correctly classified images on both original and segmented datasets. Notably, in the original images (see Fig. 11 (a) - 11 (g)), the most influential regions in the classification are not the teeth but the lips, tongue, and gums. As a result, despite the correct classification, the prediction is not truly valid. With the application of the segmentation process, irrelevant areas are masked out, allowing the model to focus on the regions relevant to caries diagnosis. For example, in Fig. 11 (h) – 11 (j), which depicts normal teeth, the model correctly focuses on the healthy tooth structure for classification. In the case of teeth with carious lesions (see Fig. 11 (k)), the most significant contributing area in the model's prediction is the region where the lesion is present. Similarly, for filled teeth, the highlighted area corresponds to the location of the dental filling (see Fig. 11 (l)). Likewise, in cases of missing teeth, the model successfully directs attention to the missing tooth area to generate accurate predictions. These visualization results demonstrate that segmentation plays a crucial role in ensuring that the model's predictions for both carious and normal teeth are valid, thereby improving its reliability.

Table 9 presents a comparison of classification accuracy between this study and previous similar works. Studies [38], [39], and [40] utilized radiographic images, primarily focusing on a single tooth, and reported accuracies ranging from 73.30% to 92%. While

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

radiographic imaging offers clearer anatomical contrast, it requires specialized equipment and exposes patients to radiation, which may limit its accessibility in certain settings. In comparison, studies [7] and [41], which employed photographic images, achieved higher accuracies (92.50% and 97%, respectively), but their datasets were more limited in scope, concentrating solely on single-tooth images. By contrast, this study employed intraoral clinical images covering five regions: labial, right buccal, left buccal, upper occlusal, and lower occlusal, offering a broader and more challenging classification scope. Additionally, this study builds upon the work by M. Fitria [8], achieving a comparable accuracy of 79.52% versus 79.17% (this study). However, a direct comparison is limited, as the study [8] applied augmentation to the entire dataset, including validation and testing sets, which may introduce bias due to the lack of purely unseen evaluation data.

**Table 9. Comparison of classification accuracy and dataset characteristics across similar studies.**

| Ref | Accuracy | Dataset |
|---|---|---|
| [39] | 73.30% | Radiographic, Single Tooth |
| [40] | 87% | |
| [41] | 92% | |
| [7] | 92.50% | Photographic, Single Tooth |
| [42] | 97% | |
| [8] | 79.52% | Photographic, Multiple Intraoral View |
| This Study | 79.17% | |

The findings of this study underscore the importance of segmentation in enhancing model focus and improving dental classification accuracy. By applying segmentation, the classification model demonstrated more refined attention to relevant dental structures, leading to better differentiation between healthy and carious teeth and enhancing model reliability. However, several limitations must be acknowledged. A significant gap between the training and validation curves across all three classification models suggests that overfitting remains a persistent challenge. While the use of UNet-based teeth segmentation has improved classification performance, the models still struggle to generalize effectively to unseen data. This issue may stem from the relatively small dataset size, which limits the model's ability to learn diverse patterns. Furthermore, the imbalance in dataset distribution, with a higher proportion of normal cases compared to carious cases, could introduce bias, making the model more likely to favor normal classifications over caries detection. Additionally, as shown in Fig. 9, the differences in classification metrics between the segmented and original datasets tend to show positive trends. However, since the confidence intervals for all metrics cross zero, the observed improvements may not be statistically significant. This outcome could be influenced by the relatively small test set and the variability in the bootstrapped estimates, which may limit the certainty of performance gains.

Future research should prioritize expanding the dataset while ensuring a balanced distribution of caries and normal cases to mitigate model bias to the majority class, reduce overfitting, and potentially enhance the statistical significance of performance improvements. In addition, methods such as class weighting or synthetic data generation can be employed to address class imbalance during model training further, helping the model to recognize underrepresented cases better and improve overall classification robustness. Beyond ensuring overall class balance, future work could also explore training and analyzing each region of the intraoral clinical images separately—such as Labial, Right Buccal, Left Buccal, Upper Occlusal, and Lower Occlusal views. By conducting region-specific training and evaluation, it becomes possible to identify whether certain areas have an unequal distribution of normal and caries cases. This approach can help detect and mitigate potential biases where the model might favor certain predictions due to the overrepresentation of a particular class in specific tooth regions. Additionally, future studies could explore more advanced deep learning architectures to improve model robustness, such as the enhanced U-Net variants like Attention U-Net or incorporating transformer-based models such as TransUNet to take advantage of their ability to capture richer contextual features. Advanced regularization techniques could also be employed to address the overfitting problem further.

The methodology presented in this study also holds potential for clinical application by integrating it into a mobile-based system, enabling more accessible and automated caries detection through deep learning. As segmentation improves the accuracy and validity of predictions by directing model focus to relevant dental regions, it strengthens the system's reliability for real-world diagnostic support. This enhancement is particularly valuable in clinical workflows, where quick, consistent, and objective assessments can assist dentists in identifying caries more effectively, especially in high-volume or resource-limited environments.

## 5. CONCLUSION

This study proposed a U-Net-based teeth segmentation technique and integrated it into a deep learning pipeline to enhance caries classification using three different architectures: ResNet-50, InceptionV3, and ResNeXt-50. The segmentation model demonstrated strong performance, with InceptionV3-UNet achieving the highest scores—an IoU of 0.9528 and a Dice score of 0.9602—indicating its effectiveness in accurately segmenting dental structures.

Segmentation application in the classification model training process consistently enhances performance

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

across all models. Each architecture achieved higher accuracy, precision, recall, and F1-score when trained on segmented images compared to the original dataset. ResNeXt-50 recorded the highest classification accuracy at 79.17%, further supporting the positive impact of segmentation on model performance. Additionally, Grad-CAM visualization results indicate that segmentation plays a crucial role in directing the classification model's focus to relevant tooth areas for caries diagnosis, thereby enhancing performance and reliability. However, confidence interval analysis indicates that despite consistent improvements across all metrics, the observed differences may not be statistically significant, as the confidence intervals of the performance differences include zero.

Future research should focus on expanding and balancing the dataset to reduce bias while exploring advanced deep learning architectures to enhance model robustness. These efforts will support the development of a more reliable diagnostic tool for automated dental caries detection.

## REFERENCES

[1] L. Cheng, L. Zhang, L. Yue, J. Ling, M. Fan, D. Yang, Z. Huang, Y. Niu, J. Liu, J. Zhao, and Y. Li, "Expert consensus on dental caries management," *International Journal of Oral Science*, vol. 14, no. 1, p. 17, 2022, doi: 10.1038/s41368-022-00167-3.

[2] A. Warreth, "Dental Caries and Its Management," *International Journal of Dentistry*, vol. 2023, pp. 1–15, Jan. 2023, doi: 10.1155/2023/9365845.

[3] N. Zubaidah, N. Dianawati, R. D. Ridwan, T. Shirakawa, K. Kuntaman, E. M. Setiawatie, M. I. Tanzil, and S. Kunarti, "The Clinical Pattern and Prevalence of Streptococcus mutans and Streptococcus sobrinus among Adult and Children Patients with Dental Caries," *Pesquisa Brasileira em Odontopediatria e Clínica Integrada*, vol. 22, p. e210117, 2022, doi: 10.1590/pboci.2022.029.

[4] J. Opydo-Szymaczek, M. Borysewicz-Lewicka, K. Andrysiak, Z. Witkowska, A. Hoffmann-Przybylska, P. Przybylski, E. Walicka, and K. Gerreth, "Clinical Consequences of Dental Caries, Parents' Perception of Child's Oral Health and Attitudes towards Dental Visits in a Population of 7-Year-Old Children," *International Journal of Environmental Research and Public Health*, vol. 18, no. 11, p. 5844, May 2021, doi: 10.3390/ijerph18115844.

[5] N. Amarasena, S. Chrisopoulos, L. M. Jamieson, and L. Luzzi, "Oral Health of Australian Adults: Distribution and Time Trends of Dental Caries, Periodontal Disease and Tooth Loss," *International Journal of Environmental Research and Public Health*, vol. 18, no. 21, p. 11539, Nov. 2021, doi: 10.3390/ijerph182111539.

[6] F. Chairunisa, A. Ramadhani, S. Takehara, K. M. Thwin, T. Z. Tun, H. Okubo, L. Hanindriyo, T. Bramantoro, and H. Ogawa, "Oral Health Status and Oral Healthcare System in Indonesia: A Narrative Review," *Journal of International Society of Preventive and Community Dentistry*, vol. 14, no. 5, pp. 352–361, Sep. 2024, doi: 10.4103/jispcd.jispcd_73_24.

[7] J. Kühnisch, O. Meyer, M. Hesenius, R. Hickel, and V. Gruhn, "Caries Detection on Intraoral Images Using Artificial Intelligence," *Journal of Dental Research*, vol. 101, no. 2, pp. 158–165, Feb. 2022, doi: 10.1177/00220345211032524.

[8] M. Fitria, M. Oktiana, H. Rahayu, K. Saddami, H. Habibie, Y. Elma, S. Janura, R. Novita, R. Putri, and M. I. Sari, "Development of Intraoral Clinical Image Dataset for Deep Learning Caries Detection," *in 2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE), Banda Aceh, Indonesia: IEEE*, Aug. 2023, pp. 194–198. doi: 10.1109/COSITE60233.2023.10249428.

[9] R. Reda, A. Zanza, A. Mazzoni, A. Cicconetti, L. Testarelli, and D. Di Nardo, "An Update of the Possible Applications of Magnetic Resonance Imaging (MRI) in Dentistry: A Literature Review," *Journal of Imaging*, vol. 7, no. 5, p. 75, Apr. 2021, doi: 10.3390/jimaging7050075.

[10] K.-C. Pentapati and H. Siddiq, "Clinical applications of intraoral camera to increase patient compliance - current perspectives," *Clinical, Cosmetic and Investigational Dentistry*, vol. Volume 11, pp. 267–278, Aug. 2019, doi: 10.2147/CCIDE.S192847.

[11] M. Fitria, Y. Elma, M. Oktiana, K. Saddami, R. Novita, R. Putri, H. Rahayu, H. Habibie, and S. Janura, "The Deep Learning Model For Decayed-Missing-Filled Teeth Detection: A Comparison Between YOLOV5 AND YOLOV8," *Jordanian Journal of Computers and Information Technology*, no. 0, p. 1, 2024, doi: 10.5455/jjcit.71-1710834785.

[12] R. Novita, R. Putri, M. Fitria, M. Oktiana, Y. Elma, H. Rahayu, S. Janura, and H. Habibie, "Performance analysis of DMF teeth detection using deep learning: A comparative study with clinical examination as quasi experimental study," *Padjadjaran Journal of Dentistry*, vol. 36, no. 1, p. 17, Mar. 2024, doi: 10.24198/pjd.vol36no1.52357.

[13] H. Henderi, T. Wahyuningsih, and E. Rahwanto, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *IJIIS: International Journal of Informatics and Information Systems*, vol. 4, no. 1, pp. 13–20, Mar. 2021, doi: 10.47738/ijiis.v4i1.73.

[14] Q. H. Nguyen, H. B. Ly, L. S. Ho, N. Al-Ansari, H. V. Le, V. Q. Tran, I. Prakash, and B. T. Pham, "Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–15, Feb. 2021, doi: 10.1155/2021/4832864.

[15] J. Sadaiyandi, P. Arumugam, A. K. Sangaiah, and C. Zhang, "Stratified Sampling-Based Deep Learning Approach to Increase Prediction Accuracy of

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

Unbalanced Dataset," *Electronics*, vol. 12, no. 21, p. 4423, Oct. 2023, doi: 10.3390/electronics12214423.
[16] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.
[17] M. Krithika Alias AnbuDevi and K. Suganthi, "Review of Semantic Segmentation of Medical Images Using Modified Architectures of UNET," *Diagnostics*, vol. 12, no. 12, p. 3064, Dec. 2022, doi: 10.3390/diagnostics12123064.
[18] R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. P. Cohen, E. Adeli, and D. Merhof, "Medical Image Segmentation Review: The Success of U-Net," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 10076-10095, Dec. 2024, doi: 10.1109/TPAMI.2024.3435571.
[19] A. Al Qurri and M. Almekkawy, "Improved UNet with Attention for Medical Image Segmentation," *Sensors*, vol. 23, no. 20, p. 8589, Oct. 2023, doi: 10.3390/s23208589.
[20] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021, doi: 10.1109/ACCESS.2021.3086020.
[21] H. Lu, Y. She, J. Tie, and S. Xu, "Half-UNet: A Simplified U-Net Architecture for Medical Image Segmentation," *Frontiers in Neuroinformatics*, vol. 16, p. 911679, Jun. 2022, doi: 10.3389/fninf.2022.911679.
[22] R. S. Aoyon, I. Hossain, M. Abdullah-Al-Wadud, and J. Uddin, "A Secured and Continuously Developing Methodology for Breast Cancer Image Segmentation via U-Net Based Architecture and Distributed Data Training," *Computer Modeling in Engineering & Sciences*, vol. 142, no. 3, pp. 2617–2640, 2025, doi: 10.32604/cmes.2025.060917.
[23] J. S. Suri, M. Bhagawati, S. Agarwal, S. Paul, A. Pandey, S. K. Gupta, L. Saba, K. I. Paraskevas, N. N. Khanna, J. R. Laird, and A. M. Johri, "UNet Deep Learning Architecture for Segmentation of Vascular and Non-Vascular Images: A Microscopic Look at UNet Components Buffered With Pruning, Explainable Artificial Intelligence, and Bias," *IEEE Access*, vol. 11, pp. 595–645, 2023, doi: 10.1109/ACCESS.2022.3232561.
[24] R. Bouali, O. Mahboub, and M. Lazaar, "Unleashing the potential of applied UNet architectures and transfer learning in teeth segmentation on panoramic radiographs," *Intelligenza Artificiale: The international journal of the AIxIA*, vol. 18, no. 2, pp. 205–217, Oct. 2024, doi: 10.3233/IA-230067.
[25] H. Alokasi and M. B. Ahmad, "The Accuracy Performance of Semantic Segmentation Network with Different Backbones," in *2022 7th International Conference on Data Science and Machine Learning Applications (CDMA)*, Riyadh, Saudi Arabia: IEEE, Mar. 2022, pp. 49–54. doi: 10.1109/CDMA54072.2022.00013.
[26] F. Van Beers, A. Lindström, E. Okafor, and M. Wiering, "Deep Neural Networks with Intersection over Union Loss for Binary Image Segmentation:," in *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*, Prague, Czech Republic: SCITEPRESS - Science and Technology Publications, 2019, pp. 438–445. doi: 10.5220/0007347504380445.
[27] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Research Notes*, vol. 15, no. 1, p. 210, 2022, doi: 10.1186/s13104-022-06096-y.
[28] Z. Krawczyk and J. Starzyński, "Segmentation of bone structures with the use of deep learning techniques," *Bulletin of the Polish Academy of Sciences Technical Sciences*, pp. 136751–136751, Mar. 2021, doi: 10.24425/bpasts.2021.136751.
[29] Z. Li, H. Zhang, Z. Li, and Z. Ren, "Residual-Attention UNet++: A Nested Residual-Attention U-Net for Medical Image Segmentation," *Applied Sciences*, vol. 12, no. 14, p. 7149, Jul. 2022, doi: 10.3390/app12147149.
[30] F. Liu, L. Gao, J. Wan, Z. L. Lyu, Y. Y. Huang, C. Liu, and M. Han, "Recognition of Digital Dental X-ray Images Using a Convolutional Neural Network," *Journal of Digital Imaging*, vol. 36, no. 1, pp. 73–79, Sep. 2022, doi: 10.1007/s10278-022-00694-9.
[31] A. Fariza, R. Asmara, E. R. Astuti, and R. H. Putra, "Tooth and Supporting Tissue Anomalies Detection from Panoramic Radiography Using Integrating Convolution Neural Network with Batch Normalization," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 2, pp. 210–222, Apr. 2024, doi: 10.22266/ijies2024.0430.19.
[32] D. Yang, C. Martinez, L. Visuña, H. Khandhar, C. Bhatt, and J. Carretero, "Detection and analysis of COVID-19 in medical images using deep learning techniques," *Scientific Reports*, vol. 11, no. 1, p. 19638, 2021, doi: 10.1038/s41598-021-99015-3.
[33] Ž. Đ. Vujovic, "Classification Model Evaluation Metrics," *Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, 2021, doi: 10.14569/IJACSA.2021.0120670.
[34] S. F. Mokhtar, Z. Md Yusof, and H. Sapiri, "Confidence Intervals by Bootstrapping Approach: A Significance Review," *Malaysian Journal of Fundamental and Applied Sciences*, vol. 19, no. 1, pp. 30–42, 2023, doi: 10.11113/mjfas.v19n1.2660.
[35] J.-C. Chien, J.-D. Lee, C.-S. Hu, and C.-T. Wu, "The Usefulness of Gradient-Weighted CAM in Assisting Medical Diagnoses," *Applied Sciences*, vol. 12, no. 15, p. 7748, Aug. 2022, doi: 10.3390/app12157748.
[36] B. N. Rao, S. K. Sabut and R. Mishra, "Deep Learning Approach to Predict Intracerebral Hemorrhage and Grad-Cam Visualization on CT Images," *2024 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence*

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

*(RAEEUCCI)*, Chennai, India, 2024, pp. 1-4, doi: 10.1109/RAEEUCCI61380.2024.10547930.
[37] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, Feb. 2020, doi: 10.1007/s11263-019-01228-7.
[38] S. Suara, A. Jha, P. Sinha, and A. A. Sekh, "Is Grad-CAM Explainable in Medical Images?," *In Proceedings of the International Conference on Computer Vision and Image Processing*, Jammu, India, Nov. 2023, pp. 124–135. doi: 10.1007/978-3-031-58181-6_11.
[39] M. Moran, M. Faria, G. Giraldi, L. Bastos, L. Oliveira, and A. Conci, "Classification of Approximal Caries in Bitewing Radiographs Using Convolutional Neural Networks," *Sensors*, vol. 21, no. 15, p. 5192, Jul. 2021, doi: 10.3390/s21155192.
[40] S. Vinayahalingam, S. Kempers, L. Limon, D. Deibel, T. Maal, M. Hanisch, S. Bergé, and T. Xi, "Classification of caries in third molars on panoramic radiographs using deep learning," *Scientific Reports*, vol. 11, no. 1, p. 12609, 2021. doi: 10.1038/s41598-021-92121-2.
[41] F. Oztekin, O. Katar, F. Sadak, M. Yildirim, H. Cakar, M. Aydogan, Z. Ozpolat, T. T. Yildirim, O. Yildirim, O. Faust, and U. R. Acharya, "An explainable deep learning model to prediction dental caries using panoramic radiograph images," *Diagnostics*, vol. 13, no. 2, p. 226, 2023. doi: 10.3390/diagnostics13020226.
[42] A. AlSayyed, A. M. Taqateq, R. Al-Sayyed, D. Suleiman, S. Shukri, E. Alhenawi, and A. M. Albsheish, "Employing CNN ensemble models in classifying dental caries using oral photographs," *International Journal of Data & Network Science*, vol. 7, no. 4, 2023. doi: 10.5267/j.ijdns.2023.8.009.

## AUTHOR BIOGRAPHY

**Muhammad Keysha Al Yassar** is a student of the Faculty of Engineering, Universitas Syiah Kuala, pursuing his Bachelor's degree in Computer Engineering Program. His research interest is in the area of Computer Vision and Artificial Intelligent including machine learning and deep learning. During his studies, he gained experience as a Programming, Software Engineering, and Signal Processing Lab Assistant. He also participated in the MBKM Bangkit Program under the Independent Study scheme with a focus on machine learning. Since 2024, he has been involved in a research group working on a Deep Learning-based Model for Caries Detection.

**Maya Fitria** began her career as a lecturer and member of the Department of Electrical Engineering and Computer Science at Syiah Kuala University in 2017 and has been there ever since. She earned her Bachelor's degree in Computer Science from the University of Indonesia (UI) in 2012. In 2013, she continued her studies at the Department of Computer Engineering, University of Duisburg-Essen, Germany, specializing in Interactive Systems and Visualization. During her studies, she received support from the DAAD-LPSDM Aceh Scholarship. She completed her master's degree in 2016, earning a Master of Science in Computer Engineering. Her research interests is in Human-Centered AI and IoT for Diagnostic and Interactive Systems.

**Maulisa Oktiana** earned her Bachelor's degree in Electrical Engineering (S.T.) from Syiah Kuala University (USK) in 2013. She obtained her Ph.D. in Electrical and Computer Engineering from Syiah Kuala University in 2020. She received a scholarship from the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia through the PMDSU scheme. In November 2018, she visited Chiba University as an exchange student. Currently, she is a lecturer at the Department of Electrical Engineering and Computer Science at Syiah Kuala University. Her research interests include image processing, biometrics, and pattern recognition.

**Muhammad Aditya Yufnanda** has been pursuing his Bachelor's degree in Computer Engineering at the Faculty of Engineering, Universitas Syiah Kuala since 2021. He has been actively involved in several organizations, competitions, and research. He serves as the Head of the Public Relations Division in the Computer Engineering Student Association. In his fifth semester, he participated in the MBKM Bangkit Program under the Independent Study scheme, with machine learning path study. Additionally, he gained experience as a Software Engineering and Multimedia Signal Processing Teaching Assistant. He also received funding for PKM competition with the title of "Internet of Things-Based Desalination System to Improve Irrigation Water Access on Coastal Agricultural Lands". Since 2024, he has been involved in a research group with research topic of Deep Learning-based Model for Caries Detection.

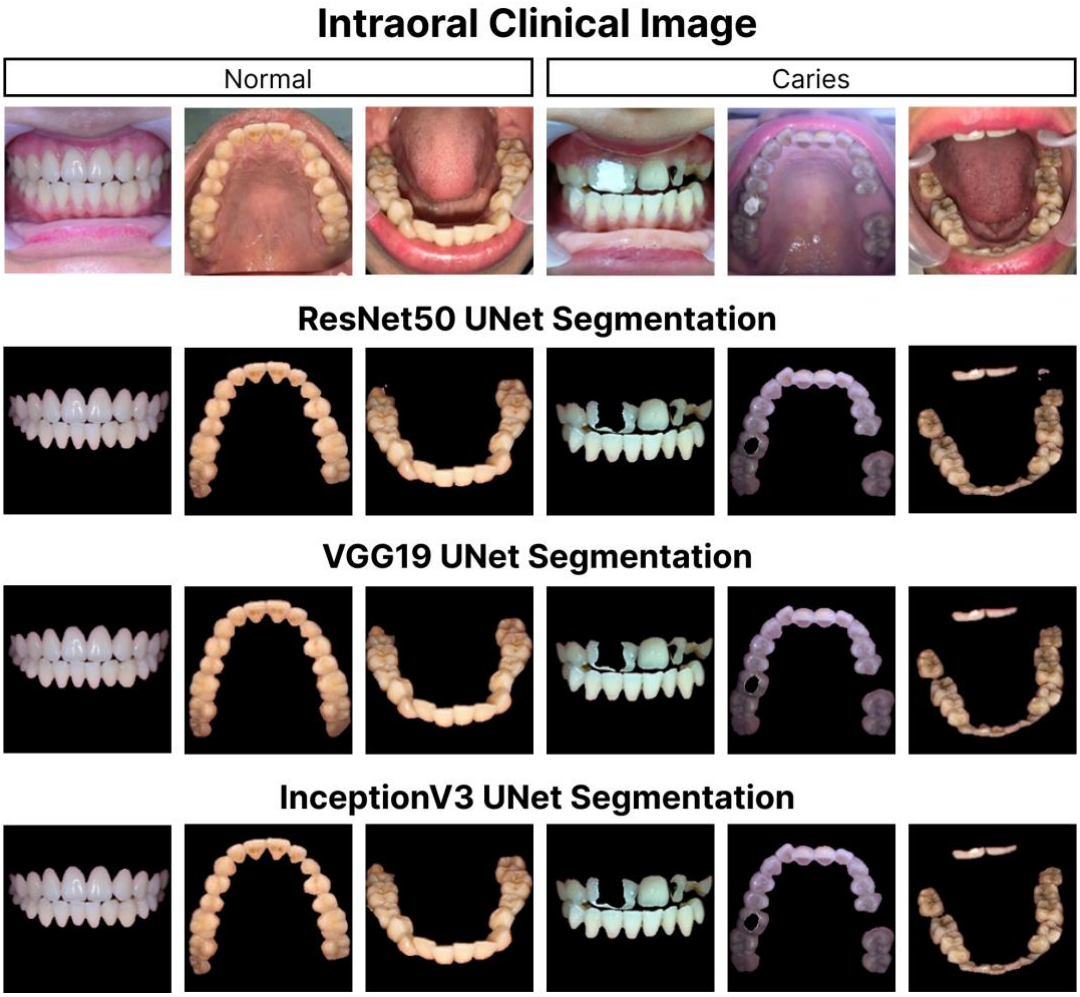
**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75

**Khairun Saddami** obtained Bachelor's degree in 2015 from Syiah Kuala University, Indonesia. He received his PhD degree in Electrical and Computer Engineering from Syiah Kuala University, Indonesia, in 2020, where he awarded a scholarship from the Ministry of Research, Technology, and Higher Education, the Republic of Indonesia, under the Scheme of Pendidikan Magister Menuju Doktor Untuk Sarjana Unggul (PMDSU). He has been with Multimedia and Signal Processing Research Group (MUSIG), Syiah Kuala University since 2020. He also acts as reviewer in reputable journal such as IEEE Access and ACM Computer Survey. He is a member of IEEE and International Association for Pattern Recognition (IAPR). His research interests are in document image analysis, deep learning, biometric and biomedical image processing and pattern recognition.

**Kahlil Muchtar** received the B.S. degree in informatics from STT-PLN (School for Engineering of PLN's Foundation), Jakarta, Indonesia, in 2007; the M.S. degree in computer science and information engineering from Asia University, Taichung, Taiwan in 2012; and Ph. D degree in electrical engineering from National Sun Yat-sen University (NSYSU), Kaohsiung City, Taiwan. His research interests include computer vision and image processing. In 2021, he served as general chair, IEEE IC-COSITE. He received the 2014 IEEE GCCE Outstanding Poster Award and IICM Taiwan 2017 The Best of Ph.D. Dissertation Award. Currently, he is an assistant professor with the Department of Electrical, and Computer Engineering, Universitas Syiah Kuala, Banda Aceh, Indonesia. From Oct 2021 until 2024, he is appointed as head of the computer engineering bachelor program. From 2018-2020, he was involved in a startup company as an AI Research Scientist at Nodeflux, Jakarta, Indonesia.

**Teuku Reza Auliandra Isma** received his Bachelor of Engineering degree in Informatics Engineering from Institut Teknologi Bandung in 2011. He was a freelance programmer for several software houses in Bandung before pursuing higher education degree. He received his Master of Science degree in Computer and Systems Engineering from Technische Universität Ilmenau in 2016. His research interests consist of artificial intelligence, system analysis and system optimization. He is now pursuing his PhD in the field of Artificial Intelligent in Institute of Automation, University of Rostock, Germany.

**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75




**Corresponding author:** Maya Fitria, mayafitria@usk.ac.id, Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Jl. Teungku Syech Abdur Rauf No. 7 Kopelma Darussalam, 23111, Banda Aceh, Indonesia.
**DOI:** https://doi.org/10.35882/ijeeemi.v7i2.75