Tooth and Supporting Tissue Anomalies Detection from Panoramic Radiography Using Integrating Convolution Neural Network with Batch Normalization

Arna Fariza[1]*   Rengga Asmara[1]   Eha Renwi Astuti[2]   Ramadhan Hardani Putra[2]

[1]*Department of Informatics and Computer Engineering, Politeknik Elektronika Negeri Surabaya, Indonesia*
[2]*Department of Dentomaxillofacial Radiology, Faculty of Dental Medicine, Universitas Airlangga, Indonesia*
* Corresponding author's Email: arna@pens.ac.id

Abstract: Abnormalities commonly encountered in dental practice include tooth and supporting tissue issues such as caries, periapical abnormalities, resorption, and impacted third molars. Panoramic radiographs are frequently used for image scanning in dentistry and oral surgery. Diagnosing dental anomalies can be time-consuming due to the complexity of the orthodontic area, potentially leading to inaccuracies. This research proposes an end-to-end automated detection of dental and supporting tissue anomalies in patients, encompassing cavities, periapical lesions, resorption, and impacted third molars. This study evaluated the effectiveness of employing various pre-trained Convolutional Neural Network architectures, including ResNet-50, ResNeXt-50 32×4d, Inception-V3, and EfficientNet-V2. To enhance model performance, a batch normalization technique was integrated into the classification layer of these pre-trained models. Data pre-processing techniques, including horizontal and vertical flips, as well as random affine transformations, were applied to augment the dataset. Additionally, an image normalization procedure was implemented before the training and prediction phases. In the evaluation on 202 images, the integrated ResNeXt-50 32x4d model with batch normalization achieved the highest accuracy, precision, recall, and F1-score of 83.663%, 81.615%, 81.271%, and 81.066%, respectively. Based on the F1-score, this model demonstrates promising predictions of tooth and supporting tissue anomalies in an imbalanced dataset.

Keywords: Convolutional neural network, Pre-trained model, Panoramic radiography, Tooth anomalies, Batch normalization.

1. Introduction

Computer vision to dentistry data promises excellent performance in detecting and classifying various restorative teeth, which are helpful for diagnosis, treatment, education, and forensics [1]. Machine learning is the most productive area of research on automatic interpretation of dental imaging [2]. Automatic image classification plays a vital role in diagnostics, as it saves time, minimizes human error rates, and improves diagnostic accuracy [3].

Panoramic radiographs are commonly used for image scanning in dentistry and by oral surgeons [4,5]. Panoramic radiography displays the maxillary and mandibular areas along with surrounding structures [5, 6]. It possesses advantages such as shorter scanning times, lower exposure doses, and more accurate measuring and estimating of the mouth area's location [6].

Dental anomalies and systemic diseases can be diagnosed from these radiographs [7]. Diagnosing dental abnormalities is potentially time-consuming due to the complexity of the orthodontic area, making it prone to inaccuracies [8]. Due to time constraints, dentists often focus more on symptoms and areas of interest [5]. The computational technique of using panoramic radiographs for anomaly detection serves as a valuable tool in dentistry [5]. The main challenges in panoramic radiography imaging include inadequate contrast, digital grain or noise, blurred boundaries, and superimposition [9].

Dental abnormalities commonly encountered in dental practice include caries, periapical lesions, resorption, and impacted third molars. Previous

studies on dental anomaly detection achieved satisfactory performance for a single abnormality. However, this study highlights the limitations of computer-aided dentistry as panoramic radiography of an individual covers a large area, including the maxilla and mandible.

Diagnostic tools based on deep learning models prove to be reliable and robust for clinical decision-making based on panoramic radiographs [10]. Before the advent of deep learning, computer-aided detection of dental anomalies (CAD) on panoramic radiographs utilized techniques such as texture analysis [11, 12] and segmentation [13]. The availability of big data panoramic radiographs has spurred research employing deep learning techniques for dental anomaly classification, rapidly enhancing performance in panoramic radiography image analysis. Numerous studies have focused on classifying dental anomalies using convolutional neural networks (CNN), including caries [14-16], impacted third molars [8, 10], periapical lesions [17], resorption [18], osteoporosis [19], cysts or tumors [20], and distinguishing between normal and abnormal cases [3].

Current studies indicate that utilizing deep learning algorithms to create CAD schematics benefits significantly from employing transfer learning. Transfer learning effectively leverages knowledge from well-established cutting-edge models and addresses issues related to limited data by incorporating external information into the model development process [21]. We found multiple research studies that applied transfer learning to dental image classification, specifically on intra-oral and X-ray images. These prior investigations, using models such as ResNet [21, 23], ResNeXt [22, 23], Inception [23], and EfficientNet [23], demonstrate significant promise in automating the classification of dental image anomalies with multiple labels and multiple classes.

As commonly observed in conventional methods, low-resolution and low-contrast panoramic radiography images can diminish accuracy and lead to incorrect classifications. While pre-trained CNNs have demonstrated their effectiveness in dental anomaly classification, enhancing classification accuracy is imperative through the utilization of custom datasets and deep transfer learning techniques. The comprehensive categorization of panoramic radiographs into three groups—those without dental abnormalities, those with single supernumerary teeth, and those with odontoma—has been explored using CNN transfer learning, specifically with AlexNet [24]. The outcomes affirm the potential of employing

deep learning for the identification of various dental abnormalities in panoramic radiographs. However, there is a demand for an efficient, rapid, and independent approach to accurately identify tooth and supporting tissue anomalies, including caries, periapical lesions, resorption, and impacted third molars.

The drawback of the ResNet and ResNeXt architectures lies in their computational complexity, particularly with deeper networks. Meanwhile, EfficientNet and Inception, designed for efficiency, still present challenges in processing panoramic radiographs effectively. The pre-trained model tends to have numerous parameters, resulting in a larger model size. Small datasets in dental imaging can limit the effectiveness of pre-trained model architectures.

Batch normalization (BatchNorm) is a widely used technique to enhance the convergence and generalization of neural networks, but its impact on accuracy can vary depending on the task [25]. BatchNorm is a normalization technique frequently employed in deep neural networks, offering benefits in improving the training and generalization of models [26]. In some cases, it may significantly enhance accuracy, while in others, the impact might be more moderate. The use of BatchNorm in the convolution output of the pre-trained model has demonstrated an overall improvement in deep neural network performance [27]. Incorporating batch normalization into the pre-trained model for panoramic radiography can alleviate some of the drawbacks by stabilizing training, facilitating better convergence, and enhancing the overall robustness and generalization of the models.

This research proposes an end-to-end automated detection of dental and supporting tissue anomalies in patients, including cavities, periapical lesions, resorption, and impacted third molars. This study evaluated the effectiveness of employing various pre-trained CNN architectures, including ResNet-50, ResNeXt-50 32×4d, Inception-V3, and EfficientNet-V2. A batch normalization technique was incorporated into the classification layer to enhance model performance. The evaluation involved a dataset of 1842 training examples and 202 test examples. Data pre-processing techniques, such as horizontal and vertical flips, and random affine transformations, were employed to augment the dataset. Furthermore, image normalization procedures were applied before the training and prediction phases.
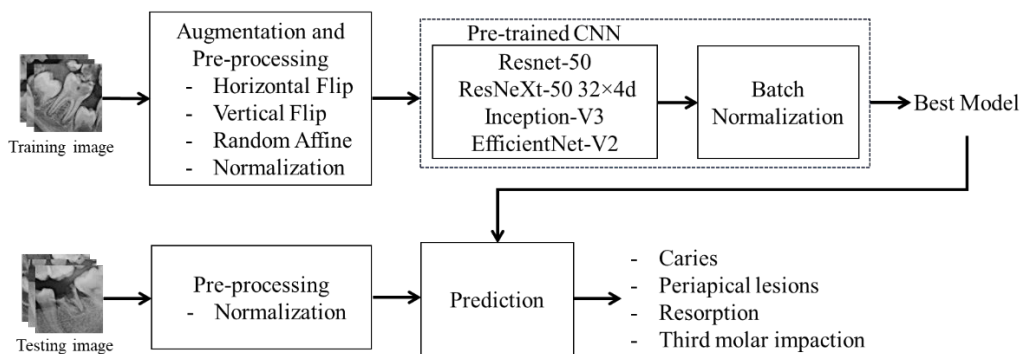
Figure. 1 Proposed methodology

2. Material and Proposed Method

The automatic detection of tooth and supporting tissue abnormalities on panoramic radiographs poses a challenge due to the heterogeneous image characteristics and low contrast. This necessitates a robust method to identify caries, periapical lesions, resorption, and impacted third molars. This research obtained ethical clearance certificate number 878/HRECC.FODM/XII/2022 from the Universitas Airlangga Faculty of Dental Medicine Health Research Ethical Clearance Commission. The proposed method in this study involves several steps, as illustrated in Fig. 1.

The input images consist of individual tooth images, spanning from the crown to the root, each measuring 224 × 224 pixels. The training and testing datasets containing abnormalities are derived from both maxillary and mandibular teeth, encompassing various tooth types (incisor, canine, premolar, and molar). Techniques such as horizontal flips, vertical flips, and random affine transformations are applied to augment the training dataset. The training dataset undergoes normalization procedures before utilizing the pre-trained CNN.

This research integrates a novel batch normalization (BatchNorm) into the classification layer of the pre-trained CNN. The best-trained model is then employed to make predictions on the testing dataset, which has also undergone normalization pre-processing.

2.1 Data collection

There are four types of dental anomalies often encountered in dentistry clinics, which are very influential on dental health: caries, periapical lesions, resorption, and impacted third molars. Dental caries is the most common chronic dental disease worldwide due to multifactors that damage hard tooth tissue [28]. Periapical periodontitis is a dental disease caused by necrotic pulp, trauma, or failed endodontic treatment. Even after endodontic treatment, the condition may progress without symptoms, eventually leading to tooth extraction [29]. Impacted third molars are the most frequently impacted teeth in humans. They can predispose the adjacent second molar to various detrimental effects such as caries, periodontitis, and cervical resorption [30]. Root resorption is a pathological process characterized by loss of tooth roots due to inflammation caused by bacterial infection, trauma, and physical or chemical irritation [18]. However, it is challenging to detect root resorption using panoramic radiography. Consequently, root resorption is often latent and undetectable.

Panoramic radiography data were collected from patients at the Parahita Diagnostic Center clinic in Sidoarjo, Indonesia. The panoramic radiography images consist of 2,044 images, each measuring 224 × 224 pixels. These images are manually cropped from each tooth on a panoramic radiography, including molars, premolars, canines, and incisors, all located in the maxilla and mandible area. They are cut in a manner that highlights the part of the tooth and supporting tissue anomalies, making it visible from the crown to the root of the tooth. It is based on the fact that dental abnormalities can occur in any part of the teeth. Fig. 2 illustrates images of teeth affected by caries, periapical lesions, resorption, and third molar impaction.

The dentist validates the construction of the dental dataset with abnormality labels. Image data processing generates the composition of the training
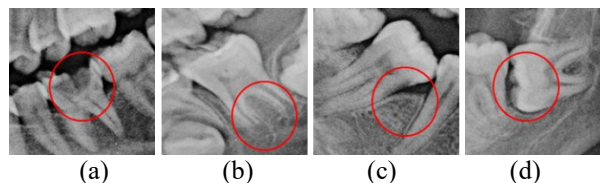

Figure. 2 Tooth and supporting tissue anomalies in 224 × 224 pixel image: (a) caries, (b) periapical lesions, (c) resorption, and (d) third molars impaction

Table 1. The composition of training and testing dataset

Tooth and Supporting Tissue Anomalies	Training Dataset Amount	Testing Dataset Amount	Total Image
Caries	385	42	427
Periapical lesions	360	39	399
Resorption	346	38	384
Third molar impaction	751	83	834
TOTAL	1,842	202	2,044

and testing datasets, as illustrated in Table 1. The training and testing datasets are randomly selected in proportion to the number of datasets per tooth abnormality, with a composition of 90% and 10%, respectively. The total number of training and testing datasets is 1,842, with 202 images each.

2.2 Augmentation and pre-processing

The augmentation process is a strategy to avoid overfitting by increasing the size of the training dataset. The training dataset, consisting of 1,842 images, was augmented to 18,420 images, including datasets for caries, periapical lesions, resorption, and third molar impaction, each with 3,850, 3,600, 3,460, and 7,510 images, respectively. The augmentation process includes horizontal flip (with a probability of 0.5), vertical flip (with a probability of 0.5), and random affine (with angles ranging from -90° to 90°). The chosen augmentation process is intended to produce different image variations while preserving the visual appearance of dental abnormalities and supporting tissues.

In addition to the augmentation process, the images in both the training and testing datasets undergo pre-processing, which involves resizing and normalization. The resizing function converts 3-channel images from their original size of 224 × 224 pixels to 3-channel images with 100 × 100 pixels dimensions. This resized format serves as the initial input for the convolutional layer of the pre-trained model. The purpose of resizing is to accelerate the image recognition process. The normalization strategy employed in pre-processing aims to standardize the data within the same range, making training more straightforward and enhancing learning speed. The pixel values are normalized using an Eq. (1) that utilizes the dataset's mean m and the standard deviation s.

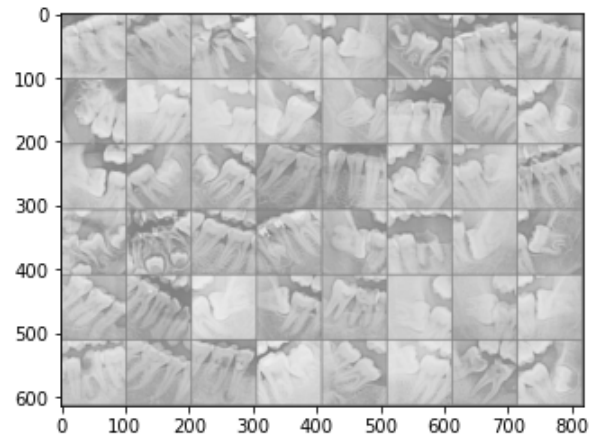$$x_{normalized} = \frac{x-m}{s} \qquad (1)$$

Figure. 3 Example of the original cropped image before augmentation and pre-processing
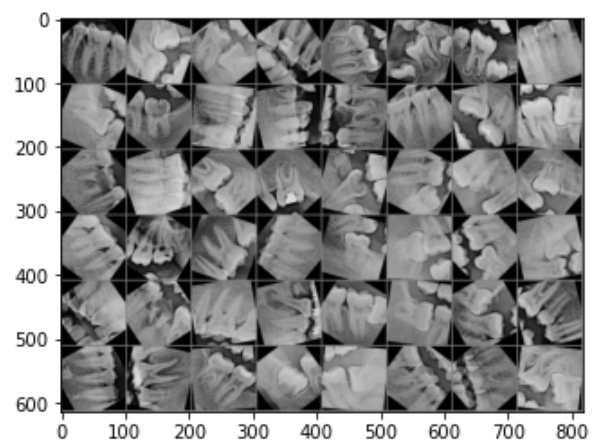

Figure. 4 Example of the cropped image after augmentation and pre-processing

The image normalization uses values $m = 0.5$ and $s = 0.5$. The choice of mean and standard deviation values is based on experimental results that visually produced the best contrast in panoramic radiography images. Fig. 3 is an example of 48 images before the augmentation and pre-processing steps, while Fig. 4 illustrates 48 images after these processes. Following augmentation, dental images exhibit a variety of alterations due to horizontal flip, vertical flip, and random affine operations. Simultaneously, normalization pre-processing proves effective in enhancing image contrast.

2.3 Proposed CNN architecture

We employ an innovative approach to classify panoramic radiography images by utilizing transfer learning with a well-known CNN architecture. In our system, we leverage transfer learning by adapting a CNN that has been previously trained on an extensive photo dataset to our specific task. This approach enhances our predictive performance and significantly reduces computational time.

Image Input = 3 × 100 × 100

↓

Pre-trained model

↓

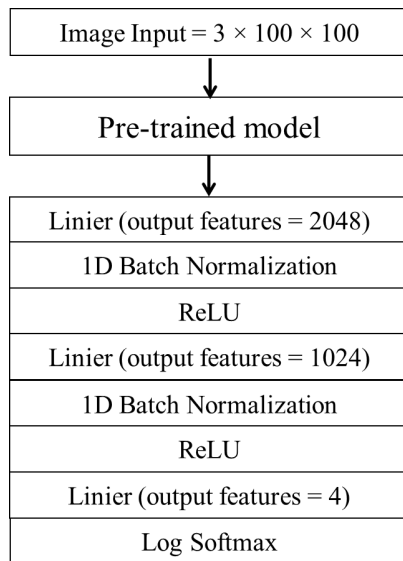| Linier (output features = 2048) |
| --- |
| 1D Batch Normalization |
| ReLU |
| Linier (output features = 1024) |
| 1D Batch Normalization |
| ReLU |
| Linier (output features = 4) |
| Log Softmax |

Figure. 5 Proposed batch normalization in the classification part of CNN pre-trained model.

The schematic representation of the CNN architecture introduced in this paper is depicted in Fig. 5. The system takes as input a 100 x 100 RGB dental image and forwards it to the CNN layer of a pre-existing model. In this research, the assessed pre-trained models included ResNet-50, ResNeXt-50 32×4d, Inception-V3, and EfficientNet-V2.

ResNet-50 is a deep convolutional neural network architecture consisting of 50 layers, specifically 48 convolutional layers, one MaxPool layer, and one average pooling layer. ResNet-50 can effectively train deep neural networks with hundreds of layers using residual connections or skip connections, reducing the problem of vanishing gradients and enabling deep network training [31]. Despite its depth, ResNet-50 uses a bottleneck design with 1x1 convolutions, reducing the number of parameters. This design improves computing efficiency while maintaining performance.

ResNeXt-50 32×4d is a neural network architecture that leverages a combination of depth and cardinality, enabling better feature representation and higher accuracy. The 32×4d configuration has multiple parallel paths, known as 'cardinalities,' for processing data [32]. These cardinalities improve resource utilization and reduce the risk of overfitting, making it an efficient choice for deep learning projects.

Inception-V3 is a convolutional neural network architecture known for its efficient and innovative design. It incorporates factorized convolutions, which reduce computational requirements by minimizing the number of parameters [33]. The architecture combines 1×1, 3×3, and 5×5

convolutional filters in parallel to effectively extract features from input images. Inception-V3 uses auxiliary classifiers as regularizers and has been recognized for its impressive performance, such as being the 1st runner-up in image classification. This architecture is well-suited for deep learning tasks, offering a balanced blend of efficiency and accuracy in computer vision applications.

EfficientNet-V2 is a neural network architecture developed by Google, known for achieving better performance with fewer parameters. It improves on the original EfficientNet by addressing architectural inefficiencies and scaling strategies [34]. The architecture is characterized by a mobile inverted bottleneck convolution similar to MobileNetV2 but significantly increased in size. EfficientNetV2 models aim to be smaller and faster while delivering state-of-the-art accuracy in image classification tasks. This compound scaling approach enables more efficient use of computational resources.

Batch Normalization (BatchNorm) is a technique initiated to normalize activations in the network, which can speed up training and improve performance [35]. This technique addresses the issue of internal covariate shift by normalizing the input layer in each mini-batch. A fully connected layer is a dense neural layer where every neuron is connected to every other neuron in an adjacent layer. BatchNorm, when applied by the fully connected layer and before the activation function, can help normalize activation, resulting in better and faster training.

In the context of a 1D fully connected neural network, BatchNorm is used to normalize the activations in the fully connected layers. It involves computing the mean (μ) and variance (σ^2) of the inputs x over a mini-batch of data. The inputs are normalized using Eq. (2).

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} \qquad (2)$$

Here, ε is a small constant added for numerical stability. Scale and shift the normalized inputs $\hat{x}$ using Eq. (3).

$$y = \gamma \hat{x} + \beta \qquad (3)$$

The γ and β are learnable parameters, allowing the network to adapt the scaling and shifting as needed. This process ensures that the inputs to the 1D fully connected layer have a standardized distribution, which helps with training stability and convergence.

The output features of the pre-trained model undergo a fully connected layer, resulting in 2048

215

neurons. Subsequently, BatchNorm and a rectified linear unit (ReLU) activation function are applied to the next layer. BatchNorm enhances the training process's speed, while ReLU is a component of a linear function that produces the input as output when it's positive and zero otherwise. BatchNorm and ReLU are applied twice, reducing the output features to 1024 and 4 neurons. A Softmax layer is added at the end of the multi-class neural network, which converts the output into a probability distribution. The softmax function transforms the output vector values to add up to 1, and it might help the training process converge quickly.

2.4 Training

The training procedure consists of two datasets, one for training and one for validation, selected randomly from a pool of 18,420 images. This selection adheres to an 80:20 ratio, resulting in 14,736 images allocated for the training dataset and 3,684 images for the validation dataset. The model's training parameters can be found in Table 2.

Categorical cross-entropy has been chosen as the loss function to be optimized. It compares the distribution of the predictions (the activations in the output layer, one for each class) to the true distribution where the probability of the true class is set to 1 and 0 for the other classes. The categorical cross-entropy loss function from a predicted output $\hat{y}_{ij}$ and true output y_{ij} for i-th class and j-th samples from N number of class and M number of samples, is given by Eq. (4).

$$L(y, \hat{y}) = -\sum_{j=0}^{M}\sum_{i=0}^{N}\left(y_{ij} \times \log(\hat{y}_{ij})\right) \qquad (4)$$

The Adaptive Moment Estimation (Adam) optimizer, is a popular optimization algorithm used in deep learning for training neural networks. Adam combines the benefits of adaptive learning rates and momentum to optimize deep neural networks efficiently and effectively. It offers faster convergence, low memory requirements, and

Table 2. Parameter for training model

Parameter	Value(s)
Learning Rate	0.0001
Optimizer	Adam (Adaptive Moment Estimation)
Loss Function	Categorical cross-entropy
Epoch	50
Batch size	30

robustness to hyperparameters, making it a preferred choice for training deep neural networks.

The training process uses a mini-batch to improve training stability and generalization performance, computational and memory efficiency, high throughput, and the ability to benefit from vectorization.

3. Result and Discussion

We performed all experiments on a computer equipped to execute the suggested model with an Intel(R) Core(TM) i7-10710U CPU running at 1.10GHz, an NVidia GeForce GTX 1650 GPU with 16.0 GB of memory, operating on a 64-bit Windows 11 system. The implementation of the proposed model was carried out using PyTorch.

Evaluating the performance of automatically identifying irregularities in teeth and the surrounding tissues on datasets with an imbalance involves assessing both the training phase and the predictive results. Following the augmentation of the training data, it comprised 18,420 images. These images were randomly divided into training and validation sets, comprising 14,736 and 3,684 images, respectively. Regarding the testing stage, 202 images were randomly chosen in a balanced manner for making predictions.

3.1 Training process evaluation

The training process is evaluated on pre-trained models ResNet-50, ResNeXt-50 32×4d, Inception-V3, and EfficientNet-V2, both without and with integrating BatchNorm. Evaluating the training process involves examining the training loss reduction and training accuracy improvement.

Fig. 6 and 9 present the trend curve of training loss using integrating BatchNorm on the pre-trained models Resnet-50 (Fig. 6), ResNeXt-50 32×4d (Fig. 7), Inception-V3 (Fig. 8), and EfficientNet-V2 (Fig. 9). The two curves show that as the training iteration period increases, the prediction accuracy of the validation subset converges to the loss rate and accuracy. The trend graph also indicates that the proposed technique does not suffer from significant overfitting or underfitting in all transfer learning models. Training with integrated BatchNorm is initially slower compared to training without it. However, over time, the performance difference diminishes, and both approaches eventually converge by the end of the training process.
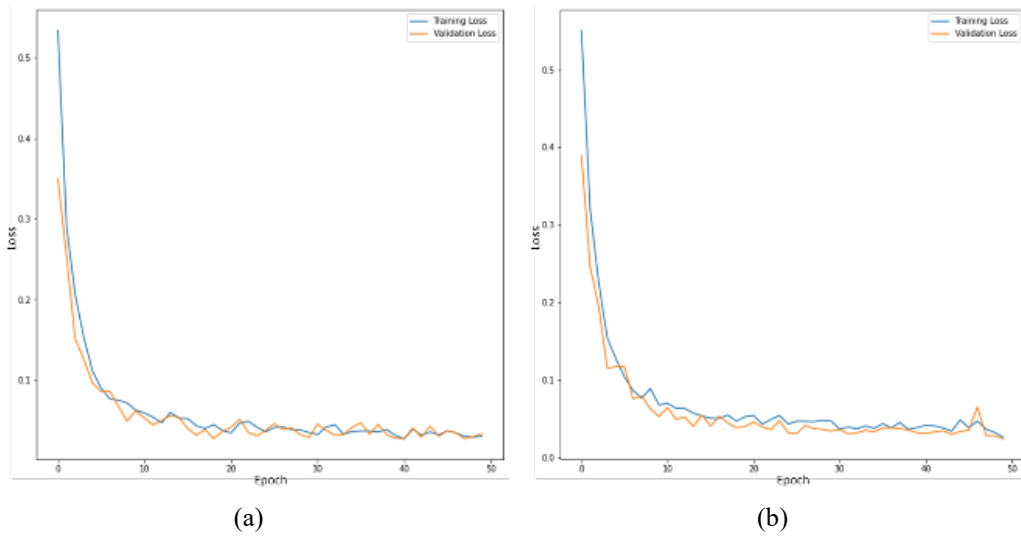
(a)   (b)

Figure. 6 Trend curve of training loss: (a) Resnet-50 and (b) Resnet-50 with BatchNorm
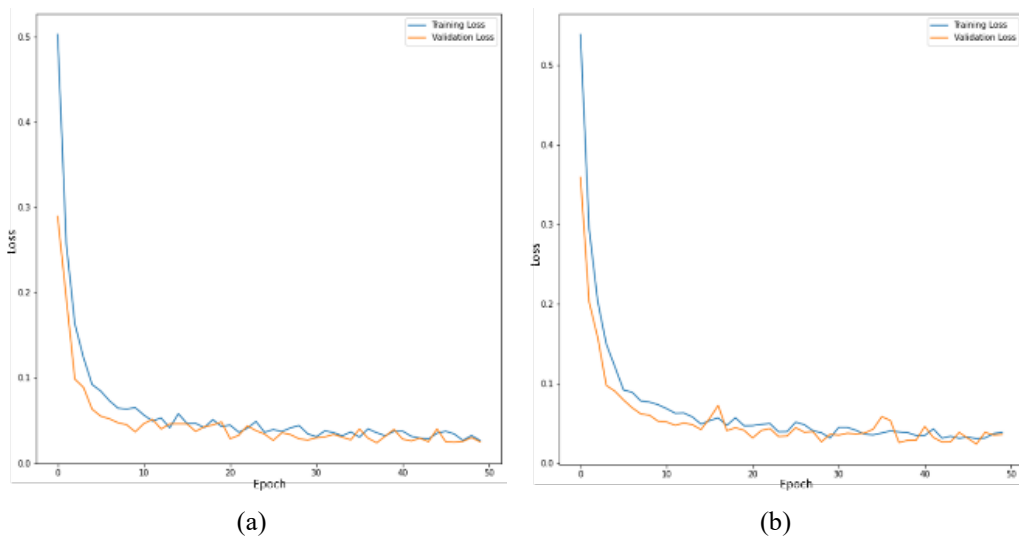

(a)   (b)

Figure. 7 Trend curve of training loss: (a) ResNeXt-50 32×4d and (b) ResNeXt-50 32×4d with BatchNorm

(a)   (b)

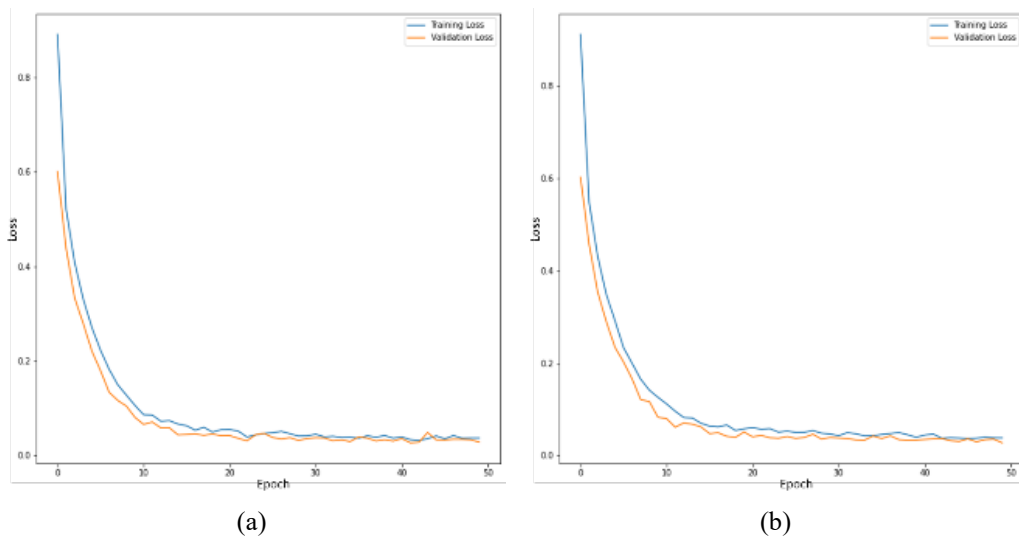Figure. 8 Trend curve of training loss: (a) Inception-V3 and (b) Inception-V3 with BatchNorm

217

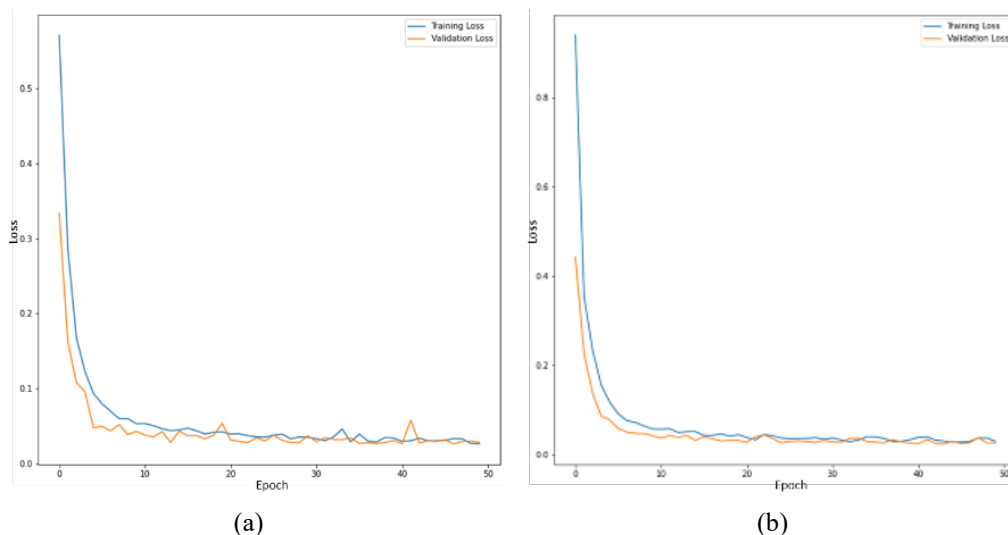(a)   (b)

Figure. 9 Trend curve of training loss: (a) EfficientNet-V2 and (b) EfficientNet-V2 with BatchNorm

Table 3. Training Performance

Architecture	Training Loss	Validation Loss	Training Accuracy (%)	Validation Accuracy (%)
Resnet-50	0.031	0.034	98.195	97.747
Resnet-50 with BatchNorm	0.026	0.024	97.991	98.317
ResNeXt-50 32×4d	0.026	0.025	98.392	98.208
ResNeXt-50 32×4d with BatchNorm	0.038	0.036	98.025	98.018
Inception-V3	0.037	0.028	97.998	98.073
Inception-V3 with BatchNorm	0.038	0.027	97.923	98.154
EfficientNet-V2	0.026	0.028	98.473	97.937
EfficientNet-V2 BatchNorm	0.030	0.027	98.181	98.127

Table 3 displays the training results following 50 epochs. The training loss for four pre-trained models, both with and without BatchNorm, achieved its minimum at 0.026 (specifically with Resnet-50 and ResNeXt-50 32×4d using BatchNorm) and its maximum at 0.038 (specifically with ResNeXt-50 32×4d and Inception-V3 with BatchNorm). Regarding training accuracy, the lowest value observed was 97.923% (in the case of Inception-V3 with BatchNorm), while the highest accuracy reached 98.392% (with ResNeXt-50 32×4d). The training process demonstrated optimal learning across all four architectural variations.

We affirm the significance of applying data augmentation to the training data. This augmentation expands the training dataset tenfold through horizontal flip, vertical flip, and random affine pre-processing techniques. The augmentation and normalization method we introduce effectively enables the optimal training of a deep learning model, even when dealing with a restricted and imbalanced dental X-ray image dataset.

3.2 Prediction result

Assessing the performance of predictions related to tooth and supporting tissue anomalies on the test dataset involves using various evaluation metrics, including accuracy, precision, recall, and F1-Score. These metrics are derived from values in the confusion matrix, specifically True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Accuracy quantifies the number of correct predictions made by the model across the entire dataset, as expressed in Eq. (5).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

Precision measures the accuracy of the model's positive predictions, as illustrated in Eq. (6).

$$Precision = \frac{TP}{TP+FP} \tag{6}$$

218

Table 3. Prediction Performance

CNN Pre-trained Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Resnet-50	78.713	75.433	74.629	74.274
Resnet-50 with BatchNorm	82.178	79.892	78.882	79.040
ResNeXt-50 32×4d	81.188	78.484	77.596	77.802
ResNeXt-50 32×4d with BatchNorm	**83.663**	**81.615**	**81.271**	**81.066**
Inception-V3	74.257	69.527	69.805	69.619
Inception-V3 with BatchNorm	81.188	78.706	78.637	78.416
EfficientNet-V2	80.693	76.965	77.220	77.017
EfficientNet-V2 with BatchNorm	80.198	77.441	76.080	75.861

Recall evaluates the model's ability to identify positive class samples in the dataset correctly.

$$Recall = \frac{TP}{TP+FN} \qquad (7)$$

F1 Score is a composite metric that considers precision (p) and recall (r) when assessing the model's accuracy. Eq. (8) outlines the calculation for F1-Score. Researchers often favor F1-Score for evaluating models on unbalanced data because accuracy may not be a reliable metric when each class in the dataset does not have an equal number of samples.

$$F1\ Score = \frac{2*p*r}{p+r} \qquad (8)$$

The image testing consists of 202 unbalanced dataset images, comprising 42 images of caries, 39 of periapical lesions, 38 of resorption, and 83 of third molar impaction. According to Table 3, the highest accuracy, precision, recall, and F1 Score values are achieved by the ResNeXt-50 32×4d model with BatchNorm, with values of 83.663%, 81.615%, 81.271%, and 81.066% respectively. The proposed integrating BatchNorm improves accuracy, precision, recall, and F1 Score for pre-trained models Resnet-50, ResNeXt-50 32×4d, and Inception-V3, but its slight same metrics for EfficientNet-V2. The most significant increases in accuracy, precision, recall, and F1 Score occur in the Inception-V3 model, with improvements of 9.3%, 13.2%, 12.7%, and 12.8% respectively. Meanwhile, EfficientNet-V2 with BatchNorm shows only a slight decrease and almost the same with EfficientNet-V2 in accuracy, precision, recall, and F1 Score, approximately 0.8%. It indicates that the proposed model is effective in enhancing the accuracy of predicting tooth and supporting tissue anomalies in panoramic radiography images for

integrated BatchNorm with pre-trained Resnet-50, ResNeXt-50 32×4d, and Inception-V3. But, EfficientNet-V2 is already effective even without proposed integrated BatchNorm.

Fig. 10 illustrates and compares the confusion matrices generated by four trained models. Fig. 10 (b) illustrates that when using Resnet-50 with BatchNorm, this model achieves detection accuracy similar to without BatchNorm (Fig. 10 (a)) for third molar impaction, but performs better in detecting caries and periapical lesions, with a decrease in performance for the resorption label. Fig. 10 (d) shows that the ResNeXt-50 32×4d model with BatchNorm results in a slight decrease in accuracy with a difference of one image for detecting third molar impaction and caries but excels in detecting periapical lesions and resorption labels compared to without BatchNorm (Fig. 10 (c)). In Fig. 10 (f), the performance of Inception-V3 with BatchNorm shows an increase in detection accuracy with a difference of one image for the third molar impaction label, a slight decrease in detection accuracy for the caries label, and a significant improvement in detecting periapical lesion and resorption labels compared to without BatchNorm (Fig. 10 (e)). Finally, Fig. 10 (h) shows the results for EfficientNet-V2 with BatchNorm, where detection accuracy decreases for the caries and third molar impaction labels and periapical lesions, but there is an improvement in the resorption label compared to without BatchNorm (Fig. 10 (g)).

Integrating BatchNorm into the proposed pre-trained model, involving augmentation and normalization pre-processing, yields higher accuracy than without BatchNorm for an unbalanced dataset. The general outcomes indicate that CNNs can be trained to attain leading-edge performance trends, both with and without BatchNorm. Integrating BatchNorm in pre-trained models holds the potential
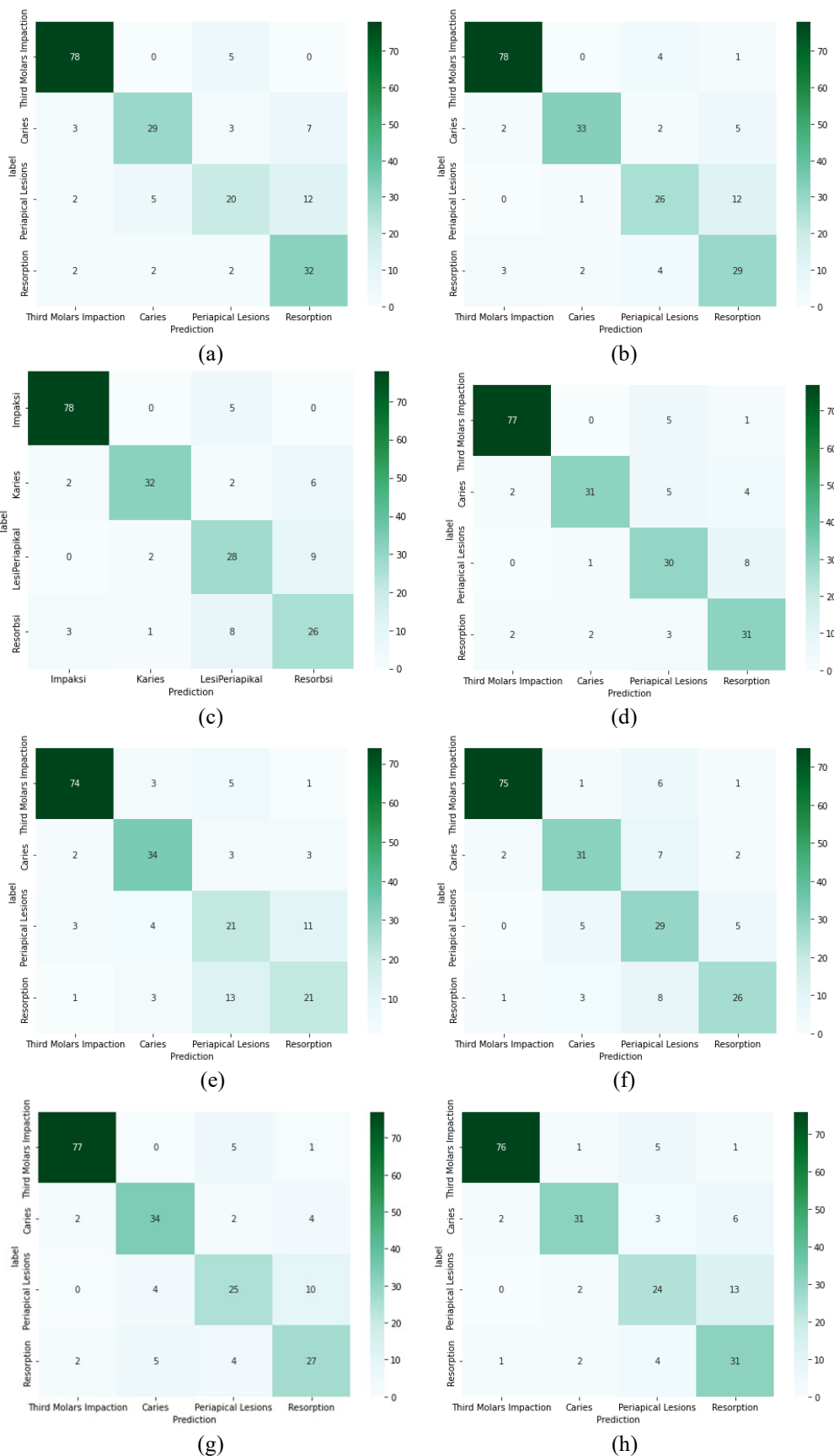
Figure. 10 Confusion matrix of prediction: (a) Resnet-50, (b) Resnet-50 with BatchNorm, (c) ResNeXt-50 32×4d, (d) ResNeXt-50 32×4d with BatchNorm, (e) Inception-V3, (f) Inception-V3 with BatchNorm, (g) EfficientNet-V2, and (h) EfficientNet-V2 with BatchNorm

for improved performance. Based on the prediction performance in Table 3 and the confusion matrix in Fig. 9, integrating the ResNeXt-50 32×4d model with BatchNorm promises the best detection of dental anomalies for panoramic radiograph images. EfficientNet, although designed to scale different resolutions efficiently, may still exhibit sensitivity to image resolution, which can be a concern in dental radiography.

Previous research by Okazaki et al. [24] used the pre-trained model AlexNet to categorize three dental anomalies, including cases without dental abnormalities, with single supernumerary teeth, and with odontoma, from panoramic radiographs. This study demonstrates a macro matrix performance of 70% for accuracy, 70.8% for precision, 70% for sensitivity, and 69.7% for the F1 score. Our proposed method achieves better accuracy, precision, sensitivity, and F1-score in classifying four dental anomalies, including caries, periapical lesions, resorption, and impacted third molars, from cropped panoramic radiography by integrating pre-trained CNNs (Resnet-50, ResNeXt-50 32×4d, Inception-V3, and EfficientNet-V2) with BatchNorm.

While the model outcomes show promise, this study does have constraints. Firstly, despite amassing and annotating a substantial dataset, the model's performance and resilience must undergo further evaluation with more extensive and diverse data due to the variety of panoramic radiography images. Secondly, to enhance the performance and robustness of the detection models, it is essential to create algorithms for image processing and classification. Future work will expand this research for more comprehensive and improved accuracy in the automatic detection of tooth and supporting tissue anomalies. Pre-processing strategies and model development remain ongoing challenges in generating more comprehensive multi-label, multi-class predictions for panoramic radiography images, even within imbalanced datasets. The employment of Computer-Aided Diagnosis (CAD) in identifying dental irregularities will serve as a valuable tool, acting as a decision support system to aid radiologist dentists in diagnosing dental anomalies.

4. Conclusion

This study uses a pre-trained CNN model to introduce a novel end-to-end approach for automatically detecting tooth and supporting tissue anomalies within an imbalanced dataset of panoramic radiographs. The BatchNorm is integrated into pre-trained CNN models alongside augmentation and normalization techniques. Evaluation of training

results on 1842 images with the proposed augmentation strategy obtained the highest accuracy on the pre-trained ResNeXt-50 32×4d model of 98,392%. Predicted on 202 images, the ResNeXt-50 32x4d model with integrating BatchNorm in the classification layer achieved accuracy, precision, recall, and F1-score of 83.663%, 81.615%, 81.271%, and 81.066%, respectively. According to the F1-score, the ResNeXt-50 32x4d model with integrating BatchNorm achieves a promising good model prediction of the tooth and supporting tissue anomalies with an imbalanced dataset.

Conflicts of Interest

The authors declare no conflict of interest.

Contributions

Conceptualization, Arna Fariza and Eha Renwi Astuti; methodology, Arna Fariza: software and validation, Arna Fariza and Rengga Asmara; formal analysis and resource, Ramadhan Hardani Putra and Eha Renwi Astuti; writing—original draft preparation, Arna Fariza; writing—review and editing, Arna Fariza, Ramadhan Hardani Putra, and Eha Renwi.

Acknowledgments

References

[1] R. Abdalla-Aslan, T. Yeshua, D. Kabla, I. Leichter, and C. Nadler, "An Artificial Intelligence System Using Machine-Learning for Automatic Detection and Classification of Dental Restorations in Panoramic Radiography", *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, Vol.130, No.5, pp.593-602, 2020.

[2] F. Pethani, "Promises and Perils of Artificial Intelligence in Dentistry", *Australian Dental Journal*, Vol.66, No.2, pp.124-135, 2021.

[3] 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: *Proc. of 2020 IEEE Region 10 Conference (TENCON)*, Osaka, Japan, pp.263-268, 2020.

[4] R. Izzetti, M. Nisi, G. Aringhieri, L. Crocetti, F. Graziani, and C. Nardi, "Basic Knowledge and

New Advances in Panoramic Radiography Imaging Techniques: A Narrative Review on What Dentists and Radiologists Should Know", *Applied Sciences,* Vol.11, No.17, pp.7858, 2021.

[5] S. Lee, D. Kim, and H. G. Jeong, "Detecting 17 Fine-Grained Dental Anomalies from Panoramic Dental Radiography Using Artificial Intelligence", *Scientific reports*, Vol.12, No.1, pp.1-8, 2022.

[6] C. Kim, D. Kim, H. Jeong, S. J. Yoon, and S. Youm, "Automatic Tooth Detection and Numbering Using a Combination of A CNN and Heuristic Algorithm", *Applied Sciences*, Vol.10, No.16, pp.5624, 2020.

[7] M. S. Kavitha, A. Asano, A. Taguchi, T. Kurita, and M. Sanada, "Diagnosis of Osteoporosis From Dental Panoramic Radiographs Using The Support Vector Machine Method in A Computer-Aided System", *BMC medical imaging*, Vol.12, No.1, pp.1-11, 2012.

[8] J. Faure, A. Engelbrecht, "Impacted Tooth Detection in Panoramic Radiographs", In: *Proc. of Advances in Computational Intelligence: 16th International Work-Conference on Artificial Neural Networks (IWANN 2021), Virtual Event, Springer International Publishing*, pp.525-536, 2021.

[9] M. Dhillon, S. Raju, S. Verma, D. Tomar, R. Mohan, M. Lakhanpal, and B. Krishnamoorthy, "Positioning Errors and Quality Assessment In Panoramic Radiography", *Imaging science in dentistry*, Vol.42, No.4. pp.207–212, 2012.

[10] M. E. Celik, "Deep Learning Based Detection Tool for Impacted Mandibular Third Molar Teeth", *Diagnostics*, Vol.12, No.4, pp.942, 2022.

[11] A. Jatti and R. Joshi, "Characterization of Dental Pathologies Using Digital Panoramic X-Ray Images Based on Texture Analysis", In: *Proc. of 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jeju, Korea, pp.592-595, 2017.

[12] M. Oda, P. V. Staziaki, M. M. Qureshi, V. C. Andreu-Arasa, B. Li, K. Takumi, and O. Sakai, "Using CT Texture Analysis to Differentiate Cystic and Cystic-Appearing Odontogenic Lesions", *European Journal of Radiology*, Vol.120, pp.108654, 2019.

[13] T. M. Tuan, H. Fujita, N. Dey, A. S. Ashour, V. T. N. Ngoc, and D. T. Chu, "Dental Diagnosis from X-Ray Images: An Expert System Based on Fuzzy Computing", *Biomedical Signal Processing and Control*, Vol.39, pp.64-73, 2018.

[14] 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, pp.1-7, 2021.

[15] T. H. Bui, K. Hamamoto, and M. P. Paing, "Automated Caries Screening Using Ensemble Deep Learning on Panoramic Radiographs", *Entropy*, Vol.24, No.10, pp. 1358, 2022.

[16] H., Mohammad-Rahimi, S. R. Motamedian, M. H. Rohban, J. Krois, S. Uribe, E. M. Nia, R. Rokhshad, M. Nadimi, and F. Schwendicke, "Deep Learning for Caries Detection: A Systematic Review: DL for Caries Detection", *Journal of Dentistry*, Vol.122, pp.104115, 2022.

[17] T. Ekert, J. Krois, L. Meinhold, K. Elhennawy, R. Emara, T. Golla, and F. Schwendicke, "Deep Learning for The Radiographic Detection of Apical Lesions", *Journal of Endodontics*, Vol.45, No.7, pp.917-922, 2019.

[18] K. Tamura, T. Kamiya, M. Oda, T. Tanaka, and Y. Morimoto, "Detection of The Root Resorption from Panoramic X-Ray Images Using Deep Metric Learning", In: *Proc. of 21st International Conference on Control, Automation and Systems (ICCAS)*, Jeju, Korea, pp.1800-1803, 2021.

[19] T. Nakamoto, A. Taguchi, and N. Kakimoto, "Osteoporosis Screening Support System from Panoramic Radiographs Using Deep Learning by Convolutional Neural Network", *Dentomaxillofacial Radiology*, Vol.51, No.6, pp.20220135, 2022.

[20] A. Lee, M. S. Kim, S. S. Han, P. Park, C. Lee, and J. P. Yun, "Deep Learning Neural Networks to Differentiate Stafne's Bone Cavity from Pathological Radiolucent Lesions of The Mandible in Heterogeneous Panoramic Radiography", *Plos one*, Vol.16, No.7, pp.e0254997, 2021.

[21] R. Ragodos, T. Wang, C. Padilla, J. T. Hecht, F. A. Poletta, F. A., I. M. Orioli, C. J. Buxó, A. Butali, C. Valencia-Ramirez, C. R. Muñeton, G. L. Wehby, S. M. Weinberg, M. L. Marazita, L. M. M. Uribe, and B. J. Howe, "Dental anomaly detection using intraoral photos via deep learning", *Scientific Reports*, Vol.12. No.1, pp.11577, 2022.

[22] A. R. Kumari, S. N. Rao, and P. R. Reddy, "Design of Hybrid Dental Caries Segmentation and Caries Detection with Meta-Heuristic-Based Resnext-RNN" *Biomedical Signal*

Processing and Control, Vol.78, pp.103961, 2022.

[23] G. Tanriver, M. Soluk Tekkesin, and O. Ergen, "Automated Detection and Classification of Oral Lesions Using Deep Learning to Detect Oral Potentially Malignant Disorders", *Cancers*, Vol.13, No.11, pp.2766, 2021.

[24] S. Okazaki, Y. Mine, Y. Iwamoto, S. Urabe, C. Mitsuhata, R. Nomura, N. Kakimoto, T. Murayama, "Analysis of the Feasibility of Using Deep Learning for Multiclass Classification of Dental Anomalies on Panoramic Radiographs", *Dental Materials Journal*, Vol.41, No.6, pp.889-895, 2022.

[25] P. Luo, X. Wang, W. Shao, Z. Peng, "Towards Understanding Regularization in Batch Normalization", *arXiv preprint arXiv:1809.00846*, 2018.

[26] E.S. Lubana, R. Dick, H. Tanaka, "Beyond Batchnorm: Towards A Unified Understanding of Normalization in Deep Learning", *Advances in Neural Information Processing Systems*, Vol.34, pp.4778-4791, 2021.

[27] Y. Zhao, J. Chen, Z. Zhang, R. Zhang, "BA-Net: Bridge attention for deep convolutional neural networks". In: *Proc of European Conference on Computer Vision*, pp. 297-312, Cham: Springer Nature Switzerland, 2022.

[28] N. J. Veiga, D. Aires, F. Douglas, M. Pereira, A. Vaz, L. Rama, M. Silva, V. Miranda, F. Pereira, B. Vidal, and J. Plaza, "Dental caries: A review", *Journal of Dental and Oral Health*, Vol.2, No.5, pp.1-3, 2016.

[29] V. T. Ngoc, D. H. Viet, L. K. Anh, D. Q. Minh, L. L. Nghia, H. K. Loan, T.M. Tuan, T.T. Ngan, and N. T. Tra, "Periapical Lesion Diagnosis Support System Based on X-ray Images Using Machine Learning Technique", *World*, Vol.12, No.3, pp.190, 2021.

[30] J. E. Enabulele and O. N. Obuekwe, "Prevalence of Caries and Cervical Resorption on Adjacent Second Molar Associated with Impacted Third Molar", Journal of Oral and Maxillofacial Surgery, *Medicine, and Pathology*, Vol.29, No.4, pp.301-305, 2017.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", In: *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.770-778, 2016.

[32] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated For Deep Neural Networks", In: *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1492-1500, 2017.

[33] C. Szegedy, V. Vanhoucke, S, Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision", In: *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.2818-2826, 2016.

[34] M. Tan and Q. Le, "Efficientnetv2: Smaller Models and Faster Training", In: *Proc. of International Conference on Machine Learning*, PMLR, pp.10096-10106, 2021.

[35] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", In: *Proc. of International Conference on Machine Learning*, PMLR, pp. 448-456, 2015.