Enhancing Dental Caries Identification with Deep Learning: A Study of Convolutional Neural Networks and Transfer Learning Approaches

Fayqa Mannan[1,2], Muhammad Adeel[3*], Erssa Arif[4], Jianqiang Li[1,5], Amin Ullah[6,7], Samra Nawazish[8] Tariq Mahmood[9, 10], Amjad Rehman[9], Ijaz Shoukat[4],

[1] Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[2] Department of Computer Science, Government College University, Faisalabad, 38000, Pakistan
[3] Department of Computer Science, School of Science, National Textile University, Faisalabad 37610, Pakistan
[4] Department of Computer Science, Faculty of Computing, Riphah International University, Faisalabad 38000, Pakistan
[5] Beijing Engineering Research Center for IoT Software and Systems, 100124, China
[6] Department of Computer Science, Faculty of IT and Computer Science, University of Central Punjab, Lahore, Pakistan
[7] School of Computer Science and Engineering Southeast University Nanjing China
[8] Software Engineering Department, University of Engineering and Technology Taxila Pakistan
[9] Artificial Intelligence & Data Analytics Lab, CCIS Prince Sultan University, Riyadh 11586, kingdom of Saudi Arabia
[10] Faculty of information Sciences, University of Education, Vehari Campus, Vehari, 61100, Pakistan

*Corresponding author: m.adeel.phd@gmail.com

Abstract: Various deep learning techniques have been employed to diagnose dental caries using X-ray images. In this study, we utilized deep learning models, including Convolutional Neural Networks (CNNs) and transfer learning models such as Visual Geometry Group (VGG16), VGG19, ResNet50, and Inception V3, to identify dental caries in periapical radiographs. The traditional CNN model comprises three sets of 2D convolutional layers, followed by an activation layer, a max-pooling layer, a flatten layer, one dense layer, another activation layer, dropout, another dense layer, and a final activation layer. For the VGG models, ResNet50, and Inception V3, the top convolutional layers were frozen to prevent retraining of the model, with only the last layers trained to produce the required output. The accuracy achieved by CNN, VGG16, VGG19, ResNet50, and Inception V3 was 90%, 96%, 73%, 70%, and 73%, respectively. VGG16 exhibited the highest accuracy among the models. Hyperparameters were carefully selected using a grid search approach, and the results were validated using the Shuffle-Split-Cross (SSC) validation method. These findings underscore the effectiveness of transfer learning in analyzing medical images, particularly in the diagnosis of dental caries.

Keywords: Deep learning; dental caries; transfer learning; X-ray images; convolutional neural network

1 Introduction

The dental caries is caused by the pathogenic bacteria. These pathogenic bacteria attack the dentin of teeth and start destroying them. The main pathogens that are responsible for dental caries are named Villanelle, Aggregatibacter, Leptotrichia, Bactericides, Granulicatella, Streptococcus, and Prevotella. Almost 80% to 90% of the global population is affected by dental caries [1]. The World Health Organization (WHO) report indicates that individuals residing in industrialized areas are primarily affected by dental caries due to unhealthy lifestyles. There is a rapid increase in sugar consumption, coupled with high exposure to fluoride [2]. Detecting all types of caries through radiography is considered one of the best methods, especially for hidden caries [3]. Dental caries are curable if diagnosed at a very early stage. In modern dentistry, the recognition of caries initiation and early detection is a primary

concern [4]. There are many traditional and convolutional methods existing for the detection of dental caries [5]. The most potent method for detecting caries is through non-invasive techniques such as the Quantitative Light-Induced Fluorescence-Digital (QLF-D) method [6]. Over the past few decades, all methods developed for caries treatment and detection have not shown significant improvement due to the morphological anatomy of teeth. While these methods are effective for detecting advanced or moderate caries, they are not useful for diagnosing early caries [7]. These methods encounter challenges in identifying deep fissures, tight inter-proximal contact, and secondary lesions on teeth, among other issues.

If caries are not detected at an early stage, they can affect not only the affected tooth but also adjacent teeth [8] [9]. It is evident that individuals living in underdeveloped countries with low to middle incomes and limited education are most affected by caries. The existing treatments are often too expensive for these populations, leading to a rapid increase in oral health issues and the potential for other systemic conditions such as cardiovascular disease, cancer, chronic respiratory disease, and diabets [10].These factors underscore the need for an algorithmic system that is both affordable and capable of detecting all types of caries at early stages, while also reducing the reliance on significant manpower and expertise.

The human tooth is mainly composed of three tissues: enamel, dentin, and pulp. Bacterial pathogens can attack any of these tissues, leading to what are respectively known as enamel caries, dentinal caries, and pulpal caries [11]. When caries develop, teeth often appear spongy, rough, decayed, broken, and exhibit varying colors. Various image processing techniques have been employed for the detection of dental caries, often followed by the use of machine learning and deep learning models [12]. Machine learning and deep learning techniques follow similar steps, including pre-processing, localization, data splitting, model training, and classification [13]. Accurately detecting dental caries at an early stage is one of the primary challenges. Recent studies have shown that deep learning algorithms have delivered outstanding performance in medical fields [14].

Our contribution involves the use of CNN and transfer learning models, specifically VGG16 and VGG19, for the detection of dental caries using periodic radiographs. These models have not been utilized for this purpose in the existing literature. The traditional CNN model comprises three sets of convolutional 2D layers, an activation layer, a max-pooling layer, and a flatten layer, followed by a dense layer, another activation layer, dropout, another dense layer, and finally an activation layer. In the VGG models, the top convolutional layers are frozen to prevent retraining of the model, and only the last layers are trained to produce the desired output.

2 Literature Review

Deep learning models are employed for preprocessing, segmentation, feature extraction, detection, and classification tasks. Lee and colleagues utilized deep convolutional neural network models to detect and diagnose dental caries using periapical radiography images. The GoogleNet Inception V3 (GNIV3) was used for data preprocessing [15]. Another study noted that dental images obtained through Computed Tomography (CT) have a limited range of properties such as sensitivity and geometric values. To address these challenges, the researchers found deep convolutional neural networks to be highly beneficial, particularly for resolution enhancement. They specifically highlighted two CNN models, U-Net and the Sub-Pixel network, in their work for enhancing image resolution. The proposed models exhibited significantly improved performance for CT images, aiding in the enhanced detection of medical attributes such as shape and size [16].

The study proposed an automated system that performs segmentation of dental images based on pixel values [17]. Furthermore, the researchers utilized a deep convolutional neural network for segmenting and labeling the affected areas of the teeth, such as those impacted by gingival issues. They employed a 3D tooth model as input, utilizing a CNN model with a two-level hierarchy. First, the teeth were labeled, and then the extracted features were fed into the neural network model. This approach achieved a remarkable 99% accuracy [18].

The literature indicates that deep learning models are widely utilized for the classification and detection of caries-infected teeth. Numerous researchers have proposed various models in different forms for classification purposes. Fully convolutional neural networks have been employed for estimating the probabilities of caries. One study proposed a model based on CNN called 'Automated Dental Red Auto-fluorescence Plaque Image Classification' for classifying QLF images. The study demonstrated the use of a CNN model for the classification of tooth and non-tooth areas by employing tooth proposals [19]. Another article focused on the automated assessment of lower third molar development from panoramic radiographs using a pilot technique. Machine Learning (ML) algorithms were utilized to train the dataset, with deep convolutional neural networks employed for classification [20].

Moreover, a machine learning approach was employed for the classification of healthy gums versus unhealthy gums. The study utilized a database of 251 Radiovisiography (RVG) dental images, containing images related to three oral diseases: caries, periapical infection, and periodontitis. The research team utilized both a CNN model and VGG16 model for disease classification, achieving 73% accuracy with the CNN model and 88% accuracy with VGG16 [21]. Additionally, a Computer-Aided Diagnosis (CAD) system was proposed for the detection of dental caries, utilizing a bitewing radiographs database consisting of 3000 images. This system was based on a Fully Connected Convolutional Neural Network (FCNN) containing over 100 layers. The model effectively classified infected and non-infected teeth [22]. In another study, panoramic X-ray images were used and classified into incisors, molars, and premolars using a CNN model, specifically a modified version of AlexNet [23]. In situations such as major disasters like bomb blasts where identification becomes incredibly challenging due to the loss of many lives, dental forensic identification tests are crucial. However, labeling a large dataset during such disasters is considered highly complicated for general dentists. Deep CNN-based models have proven to be invaluable in these scenarios, as they are utilized for the detection and classification of teeth. The study employed AlexNet for the classification of identified teeth [23].

This model achieved an accuracy of 77.4%. Deep learning models prove to be very beneficial for comparative studies. For instance, CNN models are utilized to classify periapical teeth based on their condition after treatment into three classes: 'getting better,' 'getting worse,' or 'no change,' achieving a 74% F1 score [24]. In another study, the machine learning algorithm Support Vector Machine (SVM) was used to classify caries-infected and non-infected teeth using a panoramic image database for identifying proximal dental caries. This model achieved an accuracy of 81% [25]. A Neural Network was employed for the detection of dental caries using a database of periapical radiographs. They utilized an Adaptive Neural Network for this purpose, achieving an impressive 94% accuracy [25]. Machine learning models are divided into two categories: supervised learning models and unsupervised learning models. Unsupervised learning techniques are employed for the detection of dental caries through non-standardized X-ray images [26].

Table 1. Performance comparison of the detection of dental caries techniques

Reference	Accuracy	Reference	Accuracy
[2]	51%	[19]	94%
[3]	81%,76%	[20]	81%
[7]	74%	[21]	71%
[7]	88%	[23]	88.46%
[8]	82%,74%	[24]	84%
[9]	83%	[24]	87.93%
[13]	95%	[25]	61%
[14]	95%	[26]	90%
[15]	84%	[27]	74%
[17]	74%	[28]	81%
[18]	99%	[29]	96%

Table 1 presents a performance comparison of dental caries detection techniques along with their respective accuracies. CNN demonstrates exceptional performance in the fields of computer vision and image processing. CNN models offer solutions to the aforementioned challenges. They possess the learning ability to acquire features. There are several reasons for employing CNN models:

- CNN serves as an excellent feature extractor, autonomously learning features deeply from images.
- It has the capability to exploit spatial and temporal correlation data.
- CNN can learn feature representations specific to a dataset; for example, it can identify hidden features that may be overlooked by dentists, such as proximal caries.
- CNN models automatically discover all deep or hidden features from large datasets, eliminating the need for expert intervention in image processing.

Table 2. Performance comparison of the detection of dental caries techniques

Author	Classification	Authors	Classification
[2]	F-CNN	[21]	CNN
[3]	Google Net Inception V3	[22]	ML
[4]	U-net	[23]	VGG-16
[5]	Adaptive Neural Network (ADA)	[24]	CNN
[6]	VGG16 CNN	[25]	F-CNN
[7]	Alex Net	[26]	Modified Alex Net
[8]	Mask-RCNN	[27]	CNN model
[12]	Squeeze Net, MobileNet-v2, Google Net, ResNet-18 and ResNet-50	[28]	SVM
[15]	Google Net Inception V3	[29]	VGG19-CNN
[17]	ML	[29]	Unsupervised methods
[20]	CNN		

Table 2 presents a performance comparison of dental caries detection techniques, where CNN not only demonstrates good performance in the field of dental caries but also in other domains. In dentistry, caries are detected through periapical and panoramic radiographs, as well as bitewing radiographs. Ocular caries are easily detected, but deep learning models yield impressive results in diagnosing proximal caries [27]. Deep learning models such as ANN, DNN, RNN, and CNN are also utilized in various medical fields, including the diagnosis of brain tumors, lung cancer, eye infections, digital pathology, chest and abdominal issues, musculoskeletal disorders, and dermatological diseases, among others. Deep Neural Networks (DNN) represent a class of deep learning models, employed for detecting and classifying Alzheimer's disease, as well as for segmenting brain tissues. Additionally, various ophthalmic diseases are diagnosed using deep learning models [28]. A simple CNN model is employed for diagnosing Color Fundus Imaging (CFI), which finds application in numerous ophthalmic disease-related tasks such as anatomical structure segmentation, detection of retinal abnormalities, eye disease diagnosis, and assessment of image quality [29].

3 Proposed Work

This study utilized convolutional neural network models and transfer learning models for diagnosing dental caries from periapical radiographs. The primary research steps included image acquisition, data augmentation, image preprocessing, and training and validating the models using the cross-validation method.

3.1 Image Acquisition

The process of gathering data is known as image acquisition. The database of dental periapical images was collected in 2013 and published in 2016. This dataset was obtained from the University Technology Malaysia (UTM) Health Center's dental clinic. The images were captured during regular checkups of university students at the clinic. The age group of the students ranged between 25 and 35 years old. Patients were informed about the data collection process. For this study, periapical dental X-ray images were utilized. These dental X-ray data were collected using a specialized X-ray machine. An intraoral X-ray machine, directly linked to a digital scanner and software known as "SIDEX XG," was used to produce digital periodic dental radiographs. The hardware and software utilized in the study are from a German company named Sirona. The dataset specifications include 120 images in JPEG format, with dimensions of 748x512 pixels. There are three main types of tooth radiographs: panoramic, bitewing, and periapical. The dataset used in this research consists of periapical radiographs. Periapical X-rays display all the teeth within one section, either the upper or lower jaw. These X-rays are among the most cost-effective and commonly used for diagnosing and treating dental conditions, offering clear views of the tooth roots and surrounding bone structures. This dataset is also publicly available.

3.2 Data Cleaning

Data cleaning is a straightforward process involving three steps: excluding blurred images, cropping, and flipping. All radiographs include dimension values on their sides. The next step involved flipping images to standardize all mixed images of the maxilla and mandibular jaw into the mandibular jaw orientation using a vertical flip.

3.3 Data Augmentation

The augmentation process was executed on the Keras framework using the ImageDataGenerator (IDG) function. Data was randomly generated with the following specifications: rotation range of 40 degrees, width shift range of 0.2, height shift range of 0.2, shear range of 0.2, zoom range of 0.2, horizontal flip set to true, and fill mode set to 'nearest.' Ten images were generated for each original image, resulting in a total of 1,150 training dataset images. Fig. 1 illustrates and describes the image augmentation techniques employed, including rotation, width shift, height shift, shear, and zoom range.

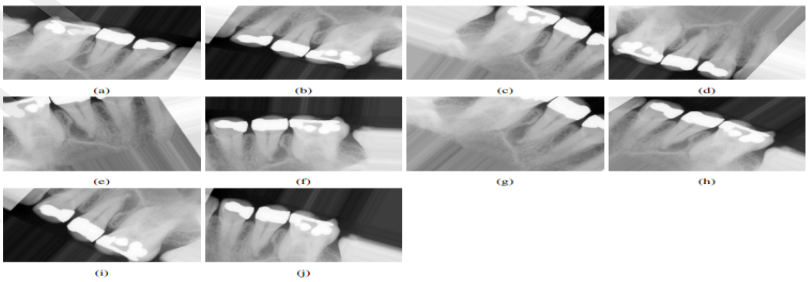


Fig. 1. Image augmentation using rotation-range, width shift & height shift-range, shear & zoom-range

3.4 Data Preprocessing

After cleaning the data, image enhancement was conducted to improve the visibility of the images. In this study, images were enhanced by adjusting three factors: brightness, sharpness (as shown in Fig. 2b), and contrast (as shown in Fig. 2c). For this purpose, the 'image enhance' function from the PILLOW library was utilized. Fig. 2 illustrates and describes the sharpening and contrast adjustments made using the 'image enhance' function, as well as the enhancement of brightness in the images.

(a) Normal Image (b) Sharpness Enhancement (c) Contrast Enhancement

Fig. 2. Image enhancement

3.5 Image Segmentation

The images were segmented using a supervised segmentation method, employing three primary techniques: thresholding algorithms, the random walker method, and the active contour method. In thresholding algorithms, images are converted into pixels, and a range of pixel values is defined by the user in supervised data. This conversion transforms the images from grayscale to binary form, where grayscale images typically contain pixel values ranging from 0 to 255. The threshold method divides the image pixels into two categories: black and white, with pixels categorized below 150 and above 150 separately using the 'threshold()' function from the OpenCV library, as depicted in Fig. 3b. Its invariance is also shown in Fig. 3c. In the active contour method, a circle or line is drawn around the Region of Interest (ROI), and this contour expands or contracts based on light and edge information. In our dataset, where multiple caries may be present in an image, we required a method that could segment all pixels from areas of small caries using predefined labels. The random walker method proved to be the most suitable for this task. The random walker method is particularly effective for segmenting a small number of pixels from images. This approach utilizes predefined pixels and 'walks' randomly to all unlabeled pixels, assigning appropriate labels. It is especially useful for high-quality image segmentation.

(a) Normal Image (b) Binary Grey Threshold (c) Binary Grey Threshold Invariance

Fig. 3. Image segmentation

Fig. 3 show the Normal image, binary grey threshold and Binary Grey threshold invariance in image segmentations.

3.6 Feature Extraction

There are various methods for extracting features from images, such as using the Principle

Component Analysis (PCA) method. PCA is utilized for the analysis of multivariate data and is particularly effective for extracting linear features and facilitating data comparisons. This method reveals only a few features by compressing the original image data. However, deep learning models are also capable of automatically extracting features from images. In this study, the models employed for detecting dental caries automatically extracted features.

3.7 Training Models

This study utilized CNN Sequential Model, VGG16, VGG19, ResNet50, and Inception V3 models. It was discovered that these models had not been previously applied to the given dataset. Details on the utilization of these models are provided below.

3.7.1 CNN Model

The input images were scaled down to 250 x 250 to reduce computational expenses. Processing high-dimensional images in their original dimensions would require significant computation. However, reducing the dimensions of images can sometimes result in the loss of information. The proposed network comprises three sets of convolutional layers, ReLU activation layers, and max-pooling layers. They are followed by a sequence of layers: flatten, dense, activation, dense, dense, and activation layer. This sequential model takes input images of size 250 x 250 x 1. Here is a summary of the inputs for all these layers:

- Input image size: 250 x 250 x 1
- Conv2D input layer size: 250 x 250 x 1
- Set of Conv2D, activation, and max-pooling layer with input size: 248 x 248 x 32
- Next set of Conv2D, activation, and max-pooling layer with input size: 124 x 124 x 32
- Set of Conv2D with input size: 61 x 61 x 32, activation: 59 x 59 x 64, and max-pooling layer with input size: 59 x 59 x 64
- Flatten layer input size: 29 x 29 x 64
- Dense layer with 53,824 nodes, followed by an activation with size 64, a dropout layer of 64, another dense layer of 64, and an activation of 1.

The kernel size is 3 x 3, all activation functions are ReLU, and the final activation function is Sigmoid. Sigmoid activation is applied to the last dense layer. The dropout layer has a size of 0.5 to mitigate the problem of overfitting.

3.7.2 Transfer Learning

VGG architectures are typically trained on ImageNet. The workings of both VGG16 and VGG19 models are the same; the only difference lies in their number of layers, with VGG19 having more layers in its structure. Firstly, in the implementation setup, the relevant packages must be imported.For VGG16 and VGG19, input images are resized, scaling them down to 224 x 224. The weights for the first three convolutional layers of VGG16 are frozen, while the subsequent layers are used for training. The concept of transfer learning is applied for fine-tuning, enabling the model to learn basic low-level features from the images in the dataset. These models are pre-loaded with ImageNet weights.

The top convolutional layers are disabled. The weights of these models are set to "ImageNet," indicating that they are pre-trained based on ImageNet. The model type is sequential, with one flattening layer included without any parameters. Three dense layers are added with units=256 and ReLU activation functions in the first and second dense layers. The last dense layer has units=3 with a "softmax" activation function. All other optional layers are excluded to make the model more adaptive to the training data. The final dense layer with 'softmax' functions as a binary classifier, distinguishing between carious and non-

carious teeth.

Inception V3:

Inception V3 architectures are typically pre-loaded with ImageNet weights. Firstly, in the implementation setup, the relevant packages must be imported. For Inception V3, input images are resized, scaling them down to 224 x 224 as shown in Fig. 4 [17]. The weights for the first three convolutional layers of Inception V3 are frozen, while the subsequent layers are used for training. The concept of transfer learning is employed for fine-tuning, enabling the model to learn basic low-level features from the dataset images. These models are pre-loaded with ImageNet weights.The top convolutional layers are disabled. The weights of these models are set to 'ImageNet,' indicating that they are pre-trained based on ImageNet. The model type is sequential, with one flattening layer included without any parameters. Three dense layers are added with units=256 and ReLU activation functions in the first and second dense layers. The last dense layer has units=3 with a 'softmax' activation function. All other optional layers are excluded to make the model more adaptive to the training data. The final dense layer with 'softmax' functions as a binary classifier, distinguishing between carious and non-carious teeth.

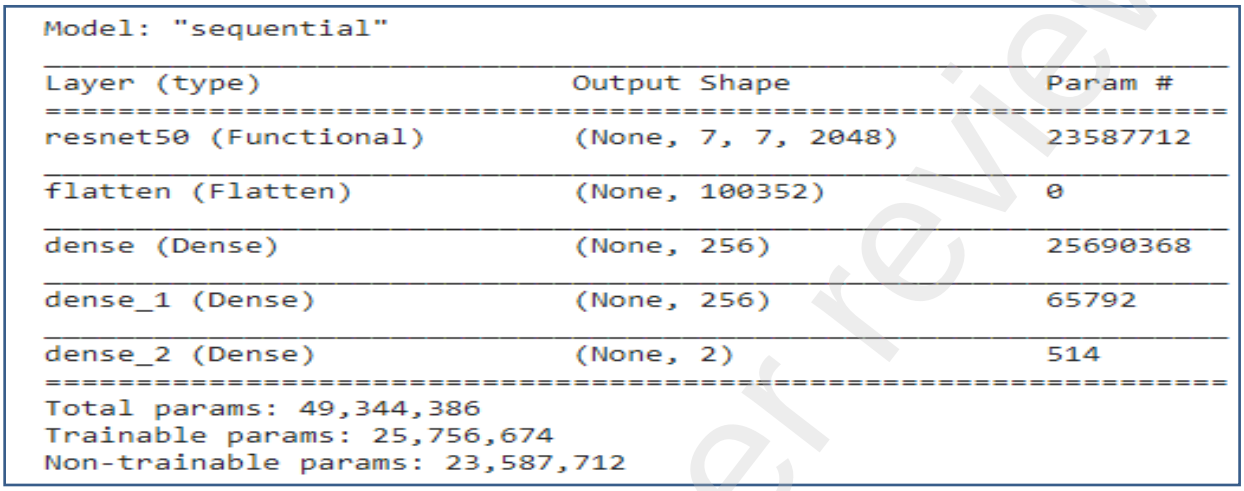
```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
inception_v3 (Functional)    (None, 5, 5, 2048)        21802784
_________________________________________________________________
flatten_1 (Flatten)          (None, 51200)             0
_________________________________________________________________
dense_3 (Dense)              (None, 256)               13107456
_________________________________________________________________
dense_4 (Dense)              (None, 256)               65792
_________________________________________________________________
dense_5 (Dense)              (None, 3)                 771
=================================================================
Total params: 34,976,803
Trainable params: 13,174,019
Non-trainable params: 21,802,784
```

Fig. 4. Inceptionv3 model used for classification of dental caries and non-infected teeth.

ResNet 50:

ResNet 50 architectures are typically pre-loaded with ImageNet weights. Firstly, in the implementation setup, the relevant packages must be imported. For ResNet 50, input images are resized, scaling them down to 224 x 224 as shown in Fig. 5 [17].The weights for the first 3 convolutional layers of ResNet 50 are frozen, while the subsequent layers are used for training. The concept of transfer learning is employed for fine-tuning, allowing the model to learn some basic low-level features from the dataset images. These models are pre-loaded with ImageNet weights. The top convolutional layers are disabled. The weights of these models are set to 'ImageNet,' indicating that they are pre-trained based on ImageNet.

In Fig. 5, it is shown that the model type is sequential, with one flatten layer included without any parameters. Additionally, three dense layers are added with units=256 and ReLU activation functions in the first and second dense layers. The last dense layer has units=3 with a 'softmax' activation function. No other optional layers are included. This is done to make the model more adaptive to the training data. The final dense layer with 'softmax' functions as a binary classifier, distinguishing between carious and non-carious teeth.



```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
resnet50 (Functional)        (None, 7, 7, 2048)        23587712
_________________________________________________________________
flatten (Flatten)            (None, 100352)            0
_________________________________________________________________
dense (Dense)                (None, 256)               25690368
_________________________________________________________________
dense_1 (Dense)              (None, 256)               65792
_________________________________________________________________
dense_2 (Dense)              (None, 2)                 514
=================================================================
Total params: 49,344,386
Trainable params: 25,756,674
Non-trainable params: 23,587,712
```

Fig. 5. ResNet 50 model used for classification of dental caries and non-infected teeth.

4 Results

All employed models efficiently classify infected teeth from non-infected ones. However, the transfer learning models show better performance than the CNN sequential model. The evaluation of these models is measured by evaluation metrics named accuracy. A summary of the models' results is provided in Table 3. In the training process of the CNN sequential model, a total of 20 epochs are needed to converge with a trained model, achieving 90% accuracy. Table 3 illustrates the training process of VGG16, VGG19, ResNet 50, and Inception V3 models, each requiring a total of 5 epochs to converge with a trained model, achieving accuracies of 96%, 73%, 70%, and 73%, respectively. The VGG16 model demonstrates 6% higher accuracy than the CNN model. Accuracy is measured as TP+TN/TP+TN+FP+FN. True Positive refers to teeth that are truly infected by dental caries and are classified as caries-infected teeth. True Negative refers to samples classified as non-infected, which are truly healthy teeth. False Positive samples are healthy but erroneously classified as caries-infected teeth. False Negative indicates teeth classified as healthy by the model, but are actually infected by caries. These terms are denoted as TP=True Positive, TN=True Negative, FP=False Positive, and FN=False Negative.

Table 3. Performance Evaluation of used models

Model	Training Accuracy	Validation Accuracy

CNN	90%	74%
VGG16	96%	96%
VGG19	73%	66%
Inception V3	70%	69%
ResNet 50	73%	66%

.

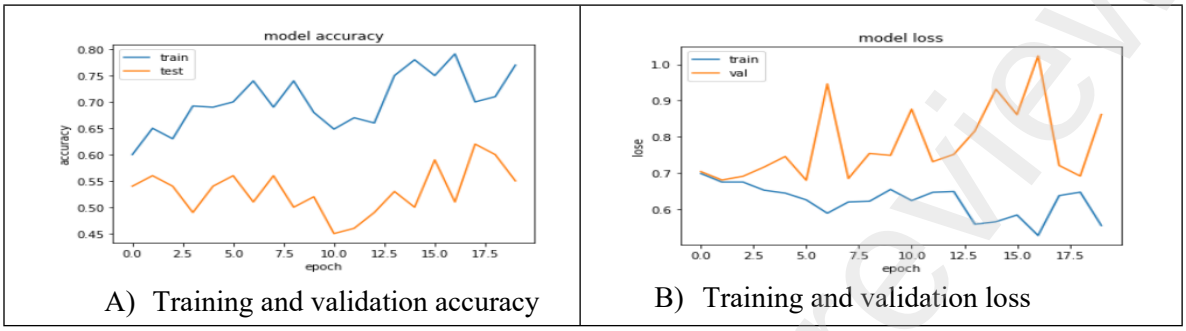
TP+TN/TP+TN+FP+FN. True Positive denotes teeth truly infected by dental caries, also referred to as caries-infected teeth. True Negative refers to samples classified as non-infected, which are genuinely healthy teeth. False Positive represents samples classified as healthy but incorrectly labeled as caries-infected teeth by the model. False Negative indicates teeth classified as healthy by the model, but in reality, they are infected by caries. These terms are denoted as: TP=True Positive, TN=True Negative, FP=False Positive, FN=False Negative.

However, these flexibilities are valuable for experimenting with large datasets. After validating the models, improvements in their performance are necessary. This is achieved by tuning parameters, which is both important and quite tricky. Finding the optimal parameter values for models is crucial, and the process can be complex. For parameter tuning, grid search is one of the best methods because it explores all possible combinations of parameters. Grid search is an approach that automatically selects the most accurate combination of parameters to assess performance validity and compare with other models. To implement grid search, GridSearchCV is imported from the sklearn.model_selection library. Hyperparameters are parameters provided to machine learning algorithms. In CNN algorithms, 'batch size' and 'epochs' are used as hyperparameters. The best combination is determined through the grid search method, as shown in Table 4.

Table 4. Performance Evaluation of used models

Models	**Input Size**	**Training Accuracy**	**Validation Accuracy**	**Training Loss**	**Validation Loss**	**Number of Epochs**	**Batch Size**
CNN sequential model	250X250	73%	53%	0.58	0.74	20	10
	200X200	68%	58%	0.67	0.67	10	5
	150X150	80%	58%	0.54	0.75	20	10
	50X50	77%	55%	0.55	0.86	20	10
Transfer Learning models	250X250	68%	50%	0.97	0.57	10	3
	250X250	70%	51%	0.55	0.47	14	2
	224X224	79%	49%	0.19	0.76	15	3
	224X224	69%	53%	0.63	0.09	10	3

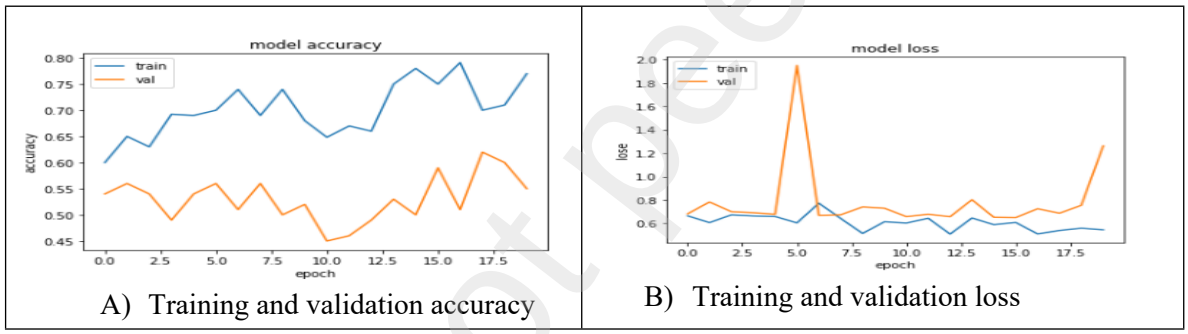
The graphs below depict the accuracy and loss of all these models. Each model was trained using varying numbers of epochs and image sizes. Included are graphs for the CNN models and VGG models. Additionally, two other models, namely ResNet 50 and Inception V3, utilized a different package but share the same architecture as the VGG models. Accuracy metrics include training and validation accuracy, as well as training and validation loss.



A) Training and validation accuracy B) Training and validation loss

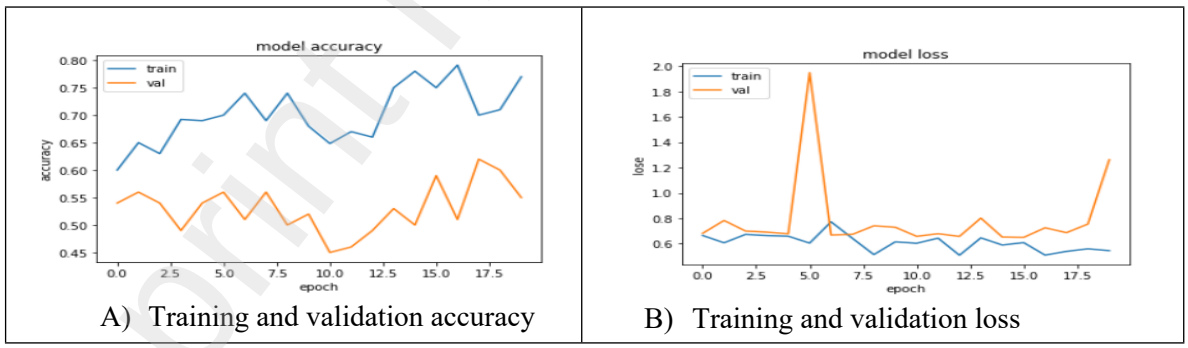
Fig. 6. CNN model graphs at the size 50*50

Fig. 6 shows the training and validation accuracy and also shown training and validation loss in CNN model graph at the size 50*50.



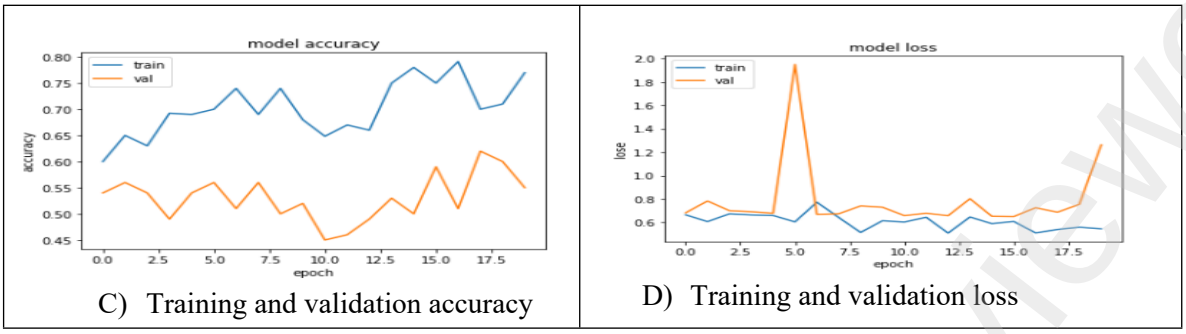
A) Training and validation accuracy B) Training and validation loss

Fig. 7. CNN model graphs at the size 150*150



A) Training and validation accuracy B) Training and validation loss

Fig. 8. CNN model graphs at the size 200*200

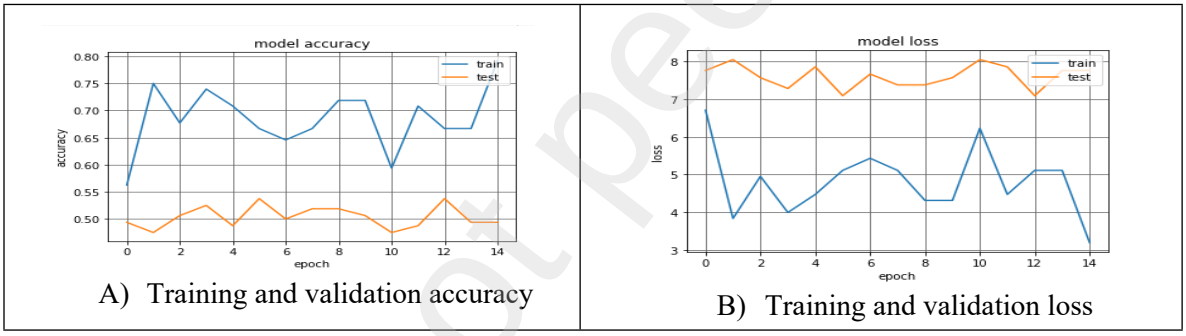


C) Training and validation accuracy

D) Training and validation loss

Fig. 9. CNN model graphs at the size 250*250

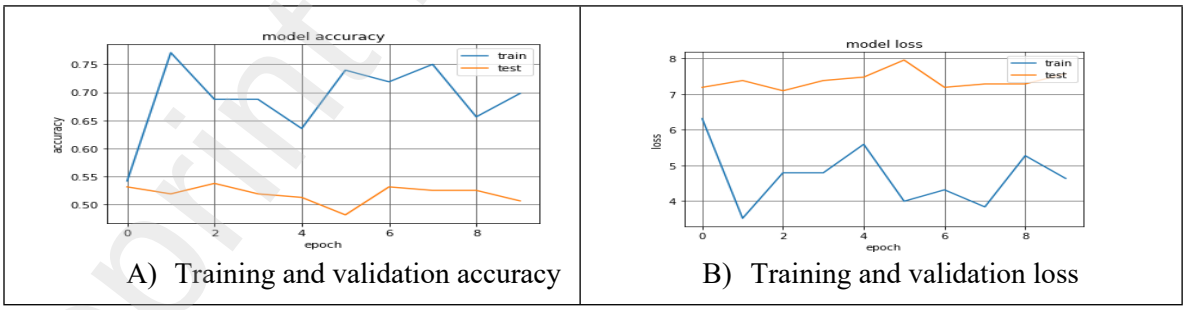
The graphs for all CNN sequential models are displayed in Fig. 7 through 13, showcasing the accuracy and loss of each model. These models were trained using varying numbers of epochs and different image sizes. The graphs illustrate the training and validation accuracy, as well as the training and validation loss, with image sizes set at 50x50, 150x150, 200x200, and 250x250 pixels. Various numbers of epochs and batch sizes are depicted, revealing the corresponding training accuracies and validation accuracies as detailed in Table 4.



A) Training and validation accuracy

B) Training and validation loss

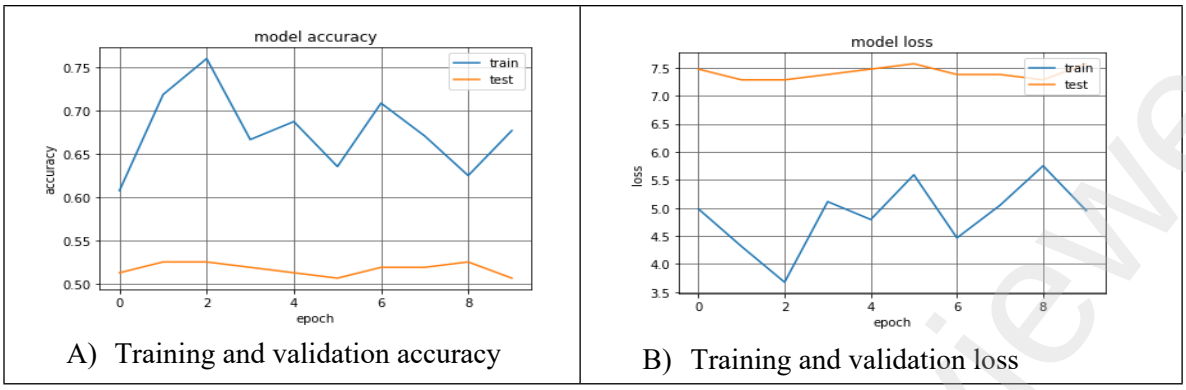
Fig. 10. VGG model graphs at the size 224*224



A) Training and validation accuracy

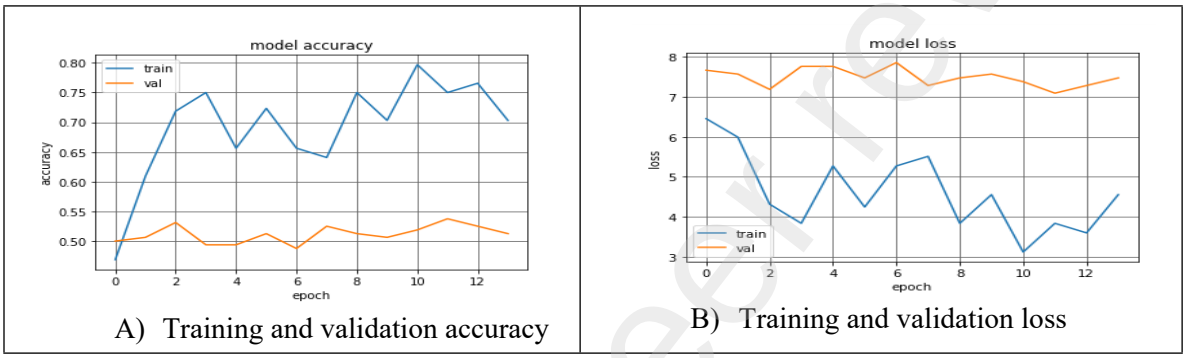
B) Training and validation loss

Fig. 11. VGG model graphs at the size 224*224



A) Training and validation accuracy B) Training and validation loss

Fig. 12. VGG model graphs at the size 224*224



A) Training and validation accuracy B) Training and validation loss

Fig. 13. VGG model graphs at the size 224*224

These graphs depict the training and validation accuracy, as well as the training and validation loss, with the input size of images set at 224x224 pixels. They illustrate various numbers of epochs and batch sizes, showcasing the varying training accuracies and validation accuracies detailed in the accompanying table.

4.1 Validation of Models

The models are validated using the cross-validation method. Among different types of cross-validation methods, the Shuffle-split-cross (SSC) validation method is particularly flexible. In this research, this method is utilized to validate the models, wherein data is split into multiple training and test sets. For this purpose, the shuffle-split function is imported from the sklearn model-selection library. This type of cross-validation offers several advantages, including the ability to control data as needed. The number of iterations can be controlled independently of training and test sizes, allowing the same part of the data to be used in each iteration by adjusting the train-size and test-size settings, which do not necessarily add up to one.However, these flexibilities prove useful when experimenting with large datasets. After validating the models, improvements in performance are necessary. This is achieved through parameter tuning, a crucial yet challenging task. To find the optimal parameter values for the models, grid search is one of the best methods. Grid search systematically explores all possible combinations of parameters, automatically selecting the most accurate combination to assess performance validity and comparison with other models. For this purpose, GridSearchCV is imported from the sklearn model-selection library. Hyperparameters, which are parameters provided to machine learning programs, play a crucial role. In CNN algorithms, "batch-size" and "epochs" are used as hyperparameters. The best combination is determined through the grid search method.

4.2 Confusion Matrix

The confusion matrix is a graphical representation commonly used in the field of neural networks for validating models. It provides a clear statistical analysis of the data, particularly in binary classification tasks, such as detecting caries or non-caries data, which is our focus. This matrix serves as a visual tool to understand the differences between true positive, true negative, false positive, and false negative outcomes. True Positive indicates teeth truly infected by dental caries, while True Negative represents samples correctly classified as non-infected, or healthy teeth. False Positive refers to samples incorrectly labeled as caries-infected when they are actually healthy, while False Negative occurs when the model misclassifies healthy teeth as infected. In our case, the confusion matrix for our test data is visually represented in Fig. 14, providing a comprehensive overview of the model's performance.

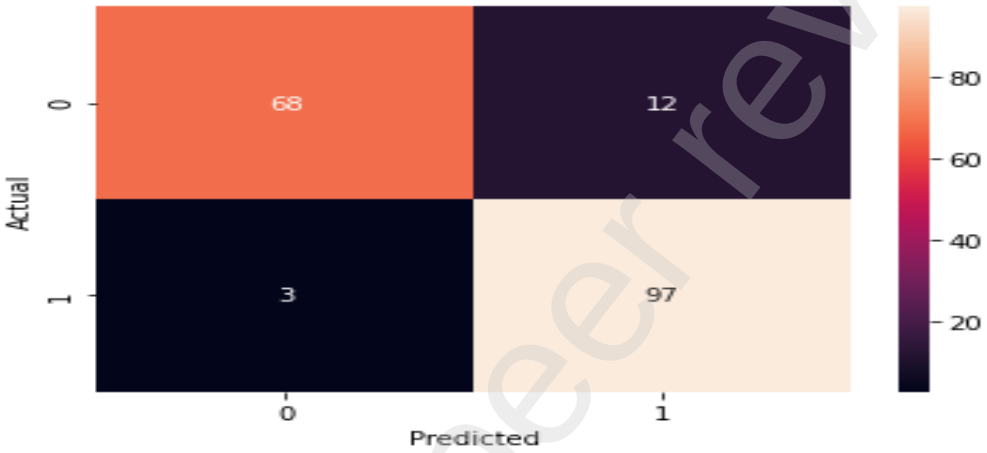


Fig. 14. Confusion matrix

4.3 ROC - Curve

The Receiver Operating Characteristic curve (ROC) is essentially a graphical representation that illustrates the performance of a model. Typically, in neural network models tasked with classifying images into binary or multiple classes, the results yield both real values and predicted values. These values serve as the basis for determining thresholds for the images. Researchers can utilize these thresholds to identify the maximum threshold value, aiming to achieve the highest possible accuracy.

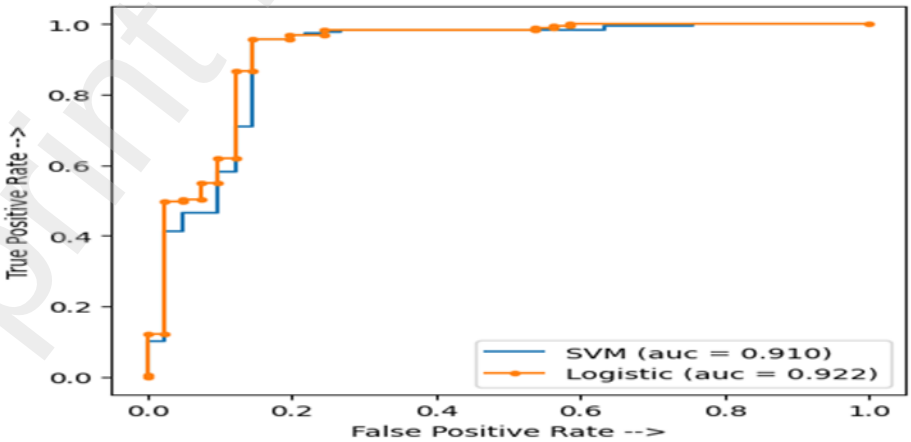


Fig. 15. ROC curve

This curve always utilizes two parameters: the actual label and the predicted label, also known as the True Positive Rate and False Positive Rate. Therefore, the ROC curve graph for the dental caries data is provided below in Fig. 15.

5 Discussion

The CNN model and transfer learning models employed for detecting dental caries demonstrate significant accuracy. Transfer learning exhibits superior accuracy compared to the CNN sequential model. The results of these models are validated through Shuffle Split-Cross validation, and hyper-parameter tuning is conducted via grid research.

5.1 Limitations

The study identifies several key limitations that need improvement for better and more accurate results. Firstly, the dataset size is notably small, which may not suffice for the convolutional neural network's requirements. Secondly, downscaled images are used as inputs to mitigate increased computational costs, training time, and storage space. Thirdly, while deep learning-based CNN methods demonstrate high accuracy and discriminatory power with high-resolution, large-scale images, this study's utilization of downscaled images may limit its potential. Fourthly, dental radiography presents challenges; grayscale images contain both light and dark regions, where distinguishing between carious regions and shadow areas can be difficult due to improper camera placement. Lastly, the type of X-ray image poses challenges, particularly periapical X-rays, which capture images from crown to root and often exhibit ambiguous boundaries between bone and teeth. Additionally, teeth in periapical images may be diversely rotated, rendering image processing more challenging.

5.2 Experimental Setup

This section discusses the experimental setup used for implementing the proposed models. It covers details about the tools, hardware, and environment utilized. The tool employed for executing both the CNN model and transfer learning models is Python, with the environment being Google Colab. Python version 3.0 is utilized, with the Keras library used on the TensorFlow backend. TensorFlow is a free and open-source software library for dataflow, primarily used for machine learning applications like neural networks. Its import is essential as it serves as the backend for the implementation setup. Keras, on the other hand, is an open-source neural network library written in Python. It provides a straightforward workflow for training and evaluating models, comprising stages such as creation, configuration, training, and evaluation of the model. For implementation, the sequential model is imported from ‘keras.models()’, while layers such as Conv2D, MaxPooling2D, Sequential, Activation, Dropout, Flatten, Dense, Input, Lambda, Flatten, etc., are imported from ‘keras.layers()’. Additionally, for creating the model, the Image Data Generator is imported from ‘keras.preprocessing.image()’.

Other libraries like NumPy, which adds support for large, multi-dimensional arrays and matrices, along with a vast collection of high-level mathematical functions to operate on these arrays, and Pandas, a software library for data manipulation and analysis in the Python programming language, are also imported. Additionally, Matplotlib, a plotting library for Python, and its numerical mathematics extension NumPy, as well as Scikit-learn (Sklearn), a machine learning library written in Python, are imported. For the implementation of transfer learning models, additional packages are required for execution. One such package, VGG16, is imported from ‘keras.application.vgg16()’. For image preprocessing, the ‘preprocess_input’ function is also imported from ‘keras.application.vgg16()’. Similarly, VGG19(), ResNet(), and InceptionV3() are imported for VGG19, ResNet50, and InceptionV3 respectively. Regarding hardware specifications needed for the implementation, a laptop with the following configuration is required: an operating system (32-bit, x64-based processor), Intel® Core™ CPU (Intel®

Core™ i3-3340 processor), CPU clocked at 2.70 GHz, Intel® integrated graphics, and 4.00GB of RAM.

6 Conclusion

Dental caries is among the most prevalent oral diseases across all age groups, underscoring the importance of early and accurate detection. The primary objective of this study is to diagnose early and proximal caries from periapical radiographs, employing CNN and transfer learning algorithms. The research is structured into several key stages: data augmentation, image preprocessing, segmentation, feature extraction, model training, and validation. Both CNN and transfer learning models are utilized for the detection of dental caries from periapical radiographs, leveraging a publicly available dataset comprising only 120 x-ray images. Transfer learning with VGG16 achieves the highest accuracy among all the models explored, reaching 96%. The experimental results of these models are discussed in detail in the results section, with validation conducted through shuffle-split-cross validation.

Acknowledgement

The authors would like to thank China's National Key R & D Program for providing the experimental facilities used to perform these experiments. The authors would also like to thank Artificial Intelligence and Data Analytics Lab CCIS Prince Sultan University for their support.

Funding Statement

This study is supported by the National Key R &D Program of China with project no. 2020YFB2104402.

Conflicts of Interest: The authors confirm that there are no conflicts of interest to disclose concerning the current study.

References

[1] O. Boiko, J. Hyttinen, P. Falt, H. berg, Mir A. Hashemi et al., "Deep learning for dental hyper spectral image analysis," In *Proceeding Color and Imaging Conference (CIC)*, Paris, France, pp. 295-299, 2019.

[2] J. Choi, H. Eun and C. Kim, "Boosting proximal dental caries detection via combination of variational methods and convolutional neural network" *Journal of Signal Processing Systems*, vol. 90, no. 1, pp. 87-97, 2018.

[3] H. J. Lee, H. D. Kim, N. S. Jeong and H. S. Cho, "Detection and diagnosis of dental caries using a deep learning based convolutional neural network algorithm," *Journal of Dentistry*, vol. 77, no. 1, pp. 106-111, 2018.

[4] G. Litjens, T. Kooi, E. B. Bejnordi, A. A. A. Setio, F. Ciompi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, no. 1, pp. 60-88, 2017.

[5] Y. Miki, C. Muramatsu, T. Hayashi, X. Zhou, T. Hara et al., "Tooth labeling in cone-beam using deep convolutional neural network for forensic identification," In *Proceeding Medical Imaging*, Orlando, Florida, United States, pp. 874-879, 2017.

[6] M. Raj Mohan, R. V. Prabhu and R. Senthil, "Diagnostic methods for early detection of dental caries a review," *International Journal of Pedodontic Rehabilitation*, vol. 1, no. 1, pp. 29, 2016.

[7] F. Casalegno, T. Newton, R. Daher, M. Abdelaziz, A. Lodi-Rizzini et al., "Caries detection with near-infrared trans-illumination using deep learning," *Journal of Dental Research*, vol. 98, no.11, pp. 1227-1233, 2019.

[8] Y. H. Ko, M. S. Kang, E. H. Kim, K. H. Kwon and I. B. Kim, "Validation of quantitative light-induced fluore science digital (qlf-d) for the detection of approximal caries in vitro," *Journal of Dentistry*, vol. 43, no. 5, pp. 568-575, 2015.

[9] E. A. Rad, M. S. M. Rahim, A. Rehman and T. Saba, "Digital dental x-ray database for caries screening," *3D Research*, vol. 7, no. 2, pp. 1-5, 2016.

[10] S. Patil, V. Kulkarni and A. Bhise, "Algorithmic analysis for dental caries detection using adaptive neural network architecture," *Heliyon*, vol. 5, no.5, pp. e01579, 2019.

[11] H. J. Lee, H. D. Kim and N. S. Jeong, "Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network" *Oral Diseases,* vol. 26, no.1, pp. 152-158, 2020.

[12] X. Xu, C. Liu and Y. Zheng, "3d tooth segmentation and labeling using deep convolutional neural networks" *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 7, pp. 2336–2348, 2018.

[13] H. J. Lee, H. D. Kim and N. S. Jeong, H. S. Choi, "Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm," *Journal of periodontal & implant science.* vol. 48, no. 2, pp. 114–123, 2018.

[14] H. M. Sultan Imangaliyev, M. C. Volgenant, B. J. Keijser, W. Crielaard and E. Levin, "Deep learning for 17 classification of dental plaque images," In *International Workshop on Machine Learning, Optimization, and Big Data*, Volterra, Italy, pp. 407-410, 2016.

[15] H. Eun and C. Kim, "Oriented tooth localization for periapical dental x-ray images via convolutional neural network," In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA),* Jeju, South Korea, pp. 1–7, 2016.

[16] J. De. Tobel, P. Radesh, D. Vandermeulen and W. P. Thevissen, "An automated technique to stage lower third molar development on panoramic radiographs for age estimation. A pilot study," *The Journal of Forensic odonto-stomatology*, vol. 35, no. 2, pp. 42, 2017.

[17] A. S. Prajapati, R. Nagaraj and S. Mitra, "Classification of dental diseases using CNN and transfer learning," In: *2017 5th International Symposium on Computational and Business Intelligence (ISCBI)*, IEEE, Dubai, United Arab Emirates, pp. 70–74, 2017.

[18] Cantu, A. Garcia, S. Gehrung, J. Krois, A. Chaurasia *et al.*, "Detecting caries lesions of different radiographic extension on bitewings using deep learning," *Journal of Dentistry*, vol. 100, no. 1, pp. 103425, 2020.

[19] A. Betul and Oktay, "Tooth detection with convolutional neural networks," In *2017 Medical Technologies National Congress (TIPTEKNO)*, Trabzon, Turkey, pp. 1–4, 2017.

[20] Naam, J. Harlan, S. Madenda and E. P. Wibowo, "Identification of the proximal caries of dental x-ray image with multiple morphology gradient method," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, no. 3, pp. 343-346, 2016.

[21] D. Osterloh and S. Viriri, "Caries detection in non-standardized periapical dental x-rays," In *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, 1st Ed, vol. 1, Springer Nature Switzerland AG, pp. 143-152, 2019.

[22] J. B. Erickson, P. Korfiatis, Z. Akkus and L. T. Kline, "Machine learning for medical imaging," *Radio-Graphics*, vol. 37, no. 2, pp. 505-515, 2017.

[23] K. Moutselos, E. Berdouses, C. Oulis and I. Maglogiannis, "Recognizing occlusal caries in dental intraoral images using deep learning," In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Berlin, Germany, pp. 1617-1620, 2019.

[24] V. D. Tuzoff, N. L. Tuzova, M. M. Bornstein, S. A. Krasnov, A. M. Kharchenko et al., "Tooth detection and numbering in panoramic radiographs using convolutional neural networks," *Den-to-maxillofacial Radiology*, vol. 48, no. 4, pp. 20180051, 2019.

[25] E. J. Kim, E. N. Nam, S. J. Shim, H. Y. Jung, H. B. Cho et al., "Transfer learning via deep neural networks for implant fixture system classification using periapical radiographs," *Journal of Clinical Medicine*, vol. 9, no. 4, pp. 1117, 2020.

[26] Y. Zhang, X. Wang, H. Li, C. Ni, Z. Du et al., "Human oral micro-biota and its modulation for oral health," *Biomedicine & Pharmacotherapy*, vol. 99, no. 1, pp. 883-893, 2018.

[27] A. Khan, A. Sohail, U. Zahoora and S. A. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 1-62, 2020.

[28] S. Vinayahalinga, T. Xi, S. Berge, T. Maal and G. De Jong, "Automated detection of third molars and mandibular nerve by deep learning," *Scientific Reports*, vol. 9, no. 1, pp. 1-7, 2019.

[29] C. Yun, L. Zhiyan, Z. Chong, L. Jing, Z. Xin et al., "Illumina-based sequencing analysis of pathogenic microorganisms in dental caries patients of different Chinese ethnic groups," *Journal of International Medical Research*, vol. 47, no. 10, pp. 5037-5047, 2019.

[30] Abdullah S. AL-Malaisa."Detection of Dental Disease through X-Ray Images using Neural Search Architecture Network". Computational Intelligence and Neuroscience 2023.

[31] Xiaotong Chen, Jiachang Guo." Detection of Proximal Caries Lesions on bitewing Radiographs using deep learning method". 2023 Feb; 56(5-6): 455-463.