International Journal of Data and Network Science 7 (2023) 1535–1550

Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Employing CNN ensemble models in classifying dental caries using oral photographs

Ayat AlSayyed[a], Abdullah Mahmoud Taqateq[a], Rizik Al-Sayyed[b*], Dima Suleiman[b], Sarah Shukri[b], Esra'a Alhenawi[c] and Ayyoub Mahmoud Albsheish[d]

[a]*School of Dentistry, The University of Jordan, Amman, Jordan*
[b]*King Abdullah II School for Information Technology, The University of Jordan, Amman, Jordan*
[c]*Department of Software Engineering, Al-Ahliyya Amman University, Amman, Jordan*
[d]*Royal Medical Services, Amman, Jordan*

C H R O N I C L E

Article history:
Received: July 2, 2023
Received in revised format: July 25, 2023
Accepted: August 10, 2023
Available online: August 10, 2023

Keywords:
Deep convolutional neural networks
Transfer learning
Ensemble learning

A B S T R A C T

Dental caries is arguably the most persistent dental condition that affects most people over their lives. Carious lesions are commonly diagnosed by dentists using clinical and visual examination along with oral radiographs. In many circumstances, dental caries is challenging to detect with photography and might be mistaken as shadows for various reasons, including poor photo quality. However, with the introduction of Artificial Intelligence and robotic systems in dentistry, photographs can be a helpful tool in oral epidemiological research for the assessment of dental caries prevalence among the population. It can be used particularly to create a new automated approach to calculate DMF (Decay, Missing, Filled) index score. In this paper, an autonomous diagnostic approach for detecting dental cavities in photos is developed using deep learning algorithms and ensemble methods. The proposed technique employs a set of pretrained models including Xception, VGG16, VGG19, and DenseNet121 to extract essential characteristics from photographs and to classify images as either normal or caries. Then, two ensemble learning methods, E- majority and E-sum, are employed based on majority voting and sum rule to boost the performances of the individual pretrained model. Experiments are conducted on 50 images with data augmentation for normal and caries images, the employed E-majority and E-sum achieved an accuracy score of 96% and 97%, respectively. The obtained results demonstrate the superiority of the proposed ensemble framework in the detection of caries. Furthermore, this framework is a step toward constructing a fully automated, efficient decision support system to be used in the dentistry area.

1. Introduction

Several techniques for image categorization have been developed in the past, ranging from classical to deep learning methods such as convolutional neural networks (CNNs). Deep learning algorithms such as Convolutional Neural Networks (CNNs) have become popular in the field of machine learning. CNNs can automatically extract the image features by adjusting the convolutional layer ("conv") and pooling layer ("pool") parameters. CNN models have been successfully used in various biomedical applications. CNNs perform classification by extracting image features directly from raw images by adjusting convolutional layer and pooling layer parameters. The features extracted by CNNs strongly depend on the size of the training dataset.

Tooth decay can be mistaken for shadows due to a variety of reasons, such as low image quality. Furthermore, many OPG images are of poor quality and do not present structures clearly due to low contrast, uneven exposure, and other reasons. In this respect, disease detection is a very challenging task and therefore Bitewing radiographs are used. Using artificial intelligence (AI) in the dentistry field, and with the introduction of CNN in dentistry, classification and the development of medical

* Corresponding author.
E-mail address: r.alsayyed@ju.edu.jo (R. Al-Sayyed)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2023.8.009

decision support systems are becoming popular topics. Oral diseases are very common ones that are widely spread. They affect people in many ways in their lifetime by causing discomfort, pain, and in severe cases might cause death. As published by the World Health Organization (WHO) 1 oral diseases affect nearly 3.5 billion people.

Dental caries is a condition in which parts of the tooth decay, which may gradually develop into small or large holes. Dental caries is one of the most prevalent health problems in different parts of the world. It is prevalent mainly among children and adolescents, but everyone may develop cavities in their mouth. If not treated, caries may grow causing severe pain, infections, and even tooth loss and other complications. Regular visits to the dentist, careful teeth cleaning, and the use of dental floss for regular and permanent dental cleaning are the best ways to prevent tooth caries.

The oral cavity contains many different types of germs where some of these germs grow and proliferate in an environment of different foods or drinks that contain cooked sugars or starches, which are also known as fermented carbohydrates. When these carbohydrates are not removed when brushing the teeth, germs may convert them into acids within 20 minutes. The germs, acids or food particles and saliva turn into dental plaque, which is a sticky layer covering the teeth. When placing the tongue on the teeth, this dental plaque can be sensed as a rough layer only a few hours after brushing the teeth, especially along the gum line.

Tooth caries is so widespread that many people do not take it seriously enough, for example, it is common not to pay attention to children's infection with caries in primary teeth, but tooth decay may lead to serious and long-term complications, even in children whose permanent teeth have not been erupted yet. Among these complications are Aches, dental abscess, chewing problems, partial or complete loss of tooth structure.

Severe dental pain can disrupt the normal practice of daily life and may cause difficulties in the chewing process which may lead to malnutrition and then weight loss. Tooth loss due to dental caries may negatively affect self-confidence. In some rare cases, a severe bacterial infection due to caries may pose a threat to the patient's life if not treated properly (Ludwig's Angina).

The dentist can diagnose tooth caries very easily. He/she asks patients if they have pain or sensitivity, then examines the mouth and teeth, and pricks the teeth with a special examination tool (a dental probe). Even more, the dentist may ask for a dental x-ray or clear oral photographs to diagnose caries. This research employs CNN ensemble models in classifying dental caries using oral photographs. Recently, much research developed caries detection models, using deep learning pretrained models.

In this research, two ensembles' methods are used to identify the caries lesions in photographs including (majority voting and sum). Four different deep learning CNNs are used in ensemble learning (Xception, VGG16, VGG19, and DenseNet121). Experiments are performed on one of the Kaggle datasets. The rest of this paper is organized as follows: Section 2 presents a recent literature review about the subject. Section 3 introduces the methodology while Section 4 describes the dataset, experiment setup, evaluation criteria, results, and discussion. Finally, Section 5 draws the conclusion and presents future work.

2. Literature Review

Deep learning pretrained models have been used in several research areas in various kinds of medical image processing, specifically dental caries detection, as they can automatically extract abstract and relevant features. Dental caries classifiers can efficiently detect the caries with accuracy that is very close to expert dentists.

Sonavane et al. (2021) proposed a caries detection model based on using the deep (CNN) The proposed model used dataset from Kaggle containing 60 images for training and 14 image for testing the model. The experimental results showed an accuracy level of 71.43%.

Haghanifar et al. (2020) used transfer learning of various pretrained deep learning models to extract relevant x-ray features, as well as a capsule network to draw predictions. An accuracy of 86.05% was achieved by the proposed model using 470 panoramic images, that include 240 labeled images for classification. Grad- As the first time of using deep learning models in detection of dental caries using panoramic x-rays, the results were encouraging and acceptable. On the other hand, the results are not remarkable when compared with previous studies that use images with better quality. In order to improve the results, two issues need to be addressed: increasing the number of panoramic images and the number of carious teeth images in the dataset. By utilizing U-Net-based segmentation models, radiologists will be able to accurately segment caries lesions.

Choudhary and his colleagues proposed an effective approach that uses VGG16 and VGG19 models with convolutional neural networks (Choudhary et al., 2021). Patients' radiographs were collected and labeled, then comparisons were made between the proposed models and other models over the collected data. Deep learning techniques can make it much easier to solve dental caries problems. Studies have shown that dental caries can be localized effectively with X-ray images. Choudhary et al. (2021) divided the research into five distinct phases: First, data collection: dental radiographs were collected from the Sharda School of Dental Sciences. dental radiographs were labeled by a team of experienced physicians. Afterward, data preprocessing was performed; preprocessing involved rotating, shifting horizontally, and widening the radiographs. Datasets are then divided into two groups: The training set (75%) and the Validation Set (25%). Their model was first trained over the training set, after that the model was validated using the validation set. To develop the dental caries classification model, pretrained deep learning models VGG16 and VGG19 were used. The final step was to evaluate the model using the following parameters (accuracy, loss, AUC, precision, and F1 score).

Stark and Samarah (2019) examined the use of ensemble and deep learning methods in real-time sensor systems. As dental caries is the most prevalent chronic disease in children, the study focused on learning methods applied to smart toothbrush devices that can improve overall health. Plaque is created when bacteria digest carbohydrates in the mouth. As a result of the acid in plaque, the teeth's hard surface is destroyed over time. Brushing removes the plaque, which prevents this from happening. Smart toothbrushes are being introduced as a solution by tracking which teeth and surfaces the user has brushed, they can show which teeth and surfaces still need brushing to ensure that plaque is removed properly. Commercial applications use a variety of techniques for this, including video recording with a smart phone or analyzing sensors data from a toothbrush with machine learning methods. The goal of this comparative study of ensemble and deep learning algorithms for real-time sensors was to evaluate which algorithm is the most promising one for detecting the tooth and surface brushed in real-time with a smart toothbrush equipped with an IMU sensor (measuring angle, velocity, angular velocity, acceleration, and position). In their study, several machine learning algorithms were evaluated as a method for predicting the localization of teeth and surfaces in real time. Data from 10 individuals were collected, pre-processed, and evaluated by cross-validation ten times. Analysis was performed on different dimensions. As a result, it was concluded that ExtraTree model produced the highest F1-score and was the best algorithm considering other aspects, such as model size and training time. Additionally, an individualized approach led to higher performance than a more generalized approach, as it is difficult to model and may perform poorly; this makes sense since brushing behavior in the observations is highly individualized.

Patients with dental caries need timely and effective treatment to reduce pain, therefore, a novel deep learning architecture was proposed by Zhu et al. (2022). The CariesNet program was used to delineate different levels of caries from panoramic radiographs. After collecting a high-resolution panoramic image; 3127 well-defined caries lesions on a panoramic radiograph, including shallow, moderate, and deep lesions caries and three types of caries were discovered. CariesNet is then built as a U-shape network with the addition of a full-scale axial attention module. In addition, segmentation performance was tested for these three caries types. In comparison with previous methods, CariesNet achieves a mean Dice coefficient of 93.6% and the segmentation of three different levels of caries was 93.61% accurate. Comparative and ablation experiments are also conducted and concluded that the new CariesNet architecture provides excellent results performance in segmenting slight lesions from large X-ray images.

Haghanifar (2022) proposed the first automated teeth extraction system for panoramic images using evolutionary algorithms. Unlike other intraoral radiography methods, Panoramic takes x-rays from outside the patient's mouth. As a result, Panoramic x-rays contain regions outside the jaw, making it extremely difficult to segment the teeth, considering that it is essential to have an image of each tooth separately to build a caries detection model, segmentation of teeth from the OPG (Orthopantomagram) image. Their proposed algorithm was applied on 42 images, where the total number of teeth was 1229; 616 maxillaries and 613 mandibular, where jaw extraction and separation were first tested, and then the efficiency of teeth extraction was investigated. In results, the achieved overall accuracy was 77.56%; where the accuracy values are 81.44% and 73.67% for maxillary and mandibular teeth, respectively. For future work, it will be easier to develop a more sophisticated neural network if there are more images with better quality and from different institutions. Efficient Net-based architecture, for example, can also be trained to increase the performance metrics and robustness as well with larger datasets.

In reference (Xue et al., 2020), authors proposed a methodology for classifying cervical histopathology pictures using ensemble transfer learning (ETL). The transfer learning structures of Xception, VGG16, VGG19, and DenseNet121 were constructed using this strategy. Based on weighted voting, an evolutionary learning technique was established. This approach was tested on a dataset of 307 photos, which were shrunk to 299 299 pixels and 224 224 pixels before being fed into the network. The highest rate of accuracy achieved was 98.61%.

Manna et al. (2021) developed an ensemble technique using InceptionV3, Xception, and DenseNet169, three pretrained CNN models [8[. A classifier fusion based on fuzzy levels was employed in their strategy. They conducted the evaluation using the SIPaKMeD and Mendeley databases. On the SIPaKMeD dataset, the accuracy rates for the 2-class and 5-class classification tasks were 98.55% and 95.43%, respectively. Using the Mendeley dataset, the accuracy percentage was 99.23.

Cunha et al. (2020) evaluated the effects of in vitro and in vivo radiation on endogenous enzymatic activity in dentin. They used a protein extracted from dentin powder of sound non-irradiated (NRT), in vitro irradiated (VTRT), and in vivo irradiated (VIRT) human teeth were used in gelatin zymographic experiments. Band densitometric analysis was used to determine their proteolytic activity. Dentin specimens were covered with fluorescein-conjugated gelatin and studied with confocal laser scanning microscopy for in-situ zymography. The hydrolyzed fluorescein-conjugated gelatin's fluorescence intensity was measured and statistically analyzed. The results showed that there was no difference in radiation treatment in vitro and in vivo. MMP-9 expression was higher in the VTRT and VIRT groups than in the NRT group. When compared to the NRT group, significant increases in gelatinolytic activity (26%for VTRT; 55%for VIRT) were seen (p 0.05).

On the other hand, Revilla-León et al. (2022) saw how the type of restorative material and its surface treatment affected the intraoral scanner's scanning accuracy, applied a mandibular dental typodont with three typodont teeth for testing, based on the crown material, ten groups were formed, an extraoral scanner (D1000; 3Shape A/G) and an intraoral scanner were used to digitize each specimen (TRIOS 4; 3Shape). The root mean square (RMS) error was calculated for each reference scan to determine the disparity with the 15 intraoral scans (Geomagic; 3D Systems). To measure trueness (0.05), they choose Welch ANOVA and GamesHowell tests. The result of their study showed that significant trueness and precision differences were found (P¡.001).

Sinha et al. (2022) used the approach of deep learning for oral disease severity detection and classification. They have conducted a survey for overall cases of oral disease being addressed in India in order to provide a backbone of deep learning for society, they have optimized convolution neural networks using bio inspired Meta heuristic algorithm (antlion and dragon fly optimizer) which give 89%accuracy by testing optimization cases of testing threshold of ROI feature extraction, and they have optimized convolution neural networks using bio inspired Meta heuristic algorithm (antlion and dragon fly They also employed a 94.6 percent accurate transfer learning model based on multilayer perceptron (MLP) for training and evaluating multiclass neural models.

Imak et al. (2022) proposed a new method for using periapical pictures to diagnose tooth caries automatically. They used a score-based ensemble technique to improve the performance of the proposed MI-DCNNE method by using a multi-input deep convolutional neural network ensemble (MI-DCNNE) model. Raw periapical pictures as well as an upgraded version were used as inputs to their suggested method. They used the Softmax layer of the proposed multi-input CNN architecture to perform the score fusion. They evaluated the performance of the proposed technique using a periapical imaging dataset (340 photos) that included both cary and non-caries images. The findings suggest that the model is extremely effective on detecting dental caries. The accuracy score is reported to be 99.13%.

Lian et al. (2021) employed pan tomographic images that were first evaluated by a team of qualified specialists. The photos depict various lesions, and a reference dataset was produced from them to serve the neural network's training data. The DenseNet had been trained to classify caries by depth and extent, whereas the nnU-Net has been trained to identify caries. The neural network's performance and accuracy were compared to the findings of three experienced dentists. The deep learning method in neural modeling had been found to be effective in detecting and classifying various lesions on panoramic radiographs, with the system's effectiveness comparable to that of a team of specialist clinicians.

3. Methodology

This research constructs an ensemble method of the pretrained models for tooth caries (TCE) diagnosis based on oral photographs. Mainly, four pretrained models (Xception (Chollet et al., 2017), DenseNet121 (Huang et al., 2017), VGG16 and VGG19 (Simonyan et al., 2014)) are utilized, then their predicted probabilities are fused to produce a final decision for a test/image. The pretrained models utilize transfer learning to mitigate these models' weights to handle a similar classification task. Ensemble learning of pretrained models attains superior performance for dentist image classification. More details are presented in Fig. 1. To boost the performance of these pretrained models, we merge their decisions using majority voting and sum rule.

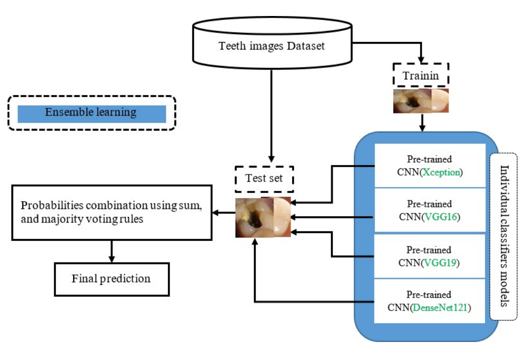


Fig. 1. The Employed Ensemble (majority voting) and Ensemble (sum)

3.1 Transfer Learning (TL) and pretrained Convolutional Neural Network Models for medical image

A machine learning method that is used to develop a model for task and then reuse the developed model as pretrained model on another task is called Transfer learning. In Transfer learning the pretrained model is trained on large dataset. There are several advantages of using Transfer learning: models that used transfer learning require less time to be trained since they are already based on pretrained model. Furthermore, the pretrained model can be used to train models with smaller datasets since it is already trained using large datasets. In this paper, four pretrained CNN models were used including Xception, VGG16, VGG19 and DenseNet121.

3.1.1 Pretrained Xception

Xception is a 71-layer pretrained convolutional neural network. In 1000 object categories, including caries and no caries, the pretrained network can categorize photos. The network has learned rich feature representations for a variety of images as a result.

3.1.2 Pretrained VGG16

VGG16 was presented by (Simonyan et al., 2014) as a deeper convolutional neural network model. The VGG16 is an object identification and classification method that has a 92.7% accuracy rate while classifying 1000 photos into 1000 different categories. It is a well-liked technique for classifying images and is simple to employ with transfer learning. VGG16 consists of 16 layers where 13 layers are convolutional, and the rest 3 layers are fully connected.

3.1.3 Pretrained VGG19

This is another pretrained convolutional neural network with 19 layers called VGG19. The trained network can identify 1000 different object types in pictures. This model consists of 19 layers, 16 layers are convolutional layers, and 3 layers are fully connected.

3.1.4 Pretrained DenseNet121

Another pretrained convolutional neural network, DenseNet121, focuses on employing shorter connections between the layers to make deep learning networks even deeper while also making them easier to train. DenseNet121 is a CNN in which every layer is connected to every layer below it. For example, the first layer is connected to the second, third, fourth, and so on layers, while the second layer is connected to the third, fourth, fifth, and so on layers. In order to maximize information flow between network tiers, this is done. DenseNet121 consists of 121 layers therefore it is divided into DenseBlocks.

3.2 Transfer Learning

This paper presents an ensemble method for dental photography images classification based on four CNN pretrained models (DenseNet (Huang et al., 2017), MobileNet (Sandler et al., 2018), VGG16 (Simonyan et al., 2014), and xception (Szegedy et al., 2016)). The output probabilities of the four pretrained models are connected to produce an 8-D feature vector (i.e., every-one with its softmax produces two probabilities based on the number of classes in dentist images). Then, various combination methods (majority voting, and sum rule) are employed to produce a final decision for the test image. Figure 1 illustrates the proposed model based on Ensemble (majority voting) and Ensemble (sum). Each pretrained model allocates an output (pre-dicted label) for the tested image.

The ensemble (majority voting) counts the votes of all the classes that belong to normal images versus the number of the classes that belong to caries images9. After that, the class that includes the maximum number of votes is nominated as the winner for this image. On the other hand, in the Ensemble(sum), the posterior probability outputs Pj(I) for each class label *j* are generated by the base classifier *t* for the test image (I). Then the class with the maximum likelihood of sum is considered the final decision. Eq. (1) shows the sum rule method in the proposed E-CNN (sum rule).

$$P(I) = \max_{j=1toc} \sum_{t=1}^{T=4} P_t^j(I) \tag{1}$$

3.3 Resources used

All the experiments are implemented using TensorFlow, Keras API, and utilized python programming in Google Colaboratory, or "CoLab." In the CoLab, we utilize Tesla GPU to run our experiment after loading the dataset into the Google drive.

4. Dataset, experiments, results, and discussion

4.1 Dataset

The experiments are conducted on dataset obtained from Kaggle. The original dataset consists of 74 teeth extracted images from panoramic x-rays. The images are distributed as follows: Caries: 55, Non-caries: 19. All images are RGB color with JPG format. Figure 2 represents a sample of the used images. We used image augmentation to increase the number of images. Thus, after augmentation the number of training set becomes (91 images) divided as follows: 46 images for non-caries and 45 images for caries. On the other hand, after augmentation the number testing set becomes (46 images) divided as follows: 24 images for non-caries and 22 images for caries.

Fig. 2. Sample of images

4.2. Experiment Setup and resources

Before we proceed, we would like to introduce the concept of image augmentation. The technique of developing new training examples from existing ones is known as image augmentation. You alter the original image slightly to create a new sample. For example, you could make a new image that is a little brighter; cut a section from the original image; 3reflect the original image, and so on.

To achieve decent results and avoid overfitting, deep neural networks require a large amount of training data. However, obtaining sufficient training samples is frequently problematic. Gathering enough data could be difficult, if not impossible, for a variety of reasons. Firstly, you must initially gather photographs and label them to create a training dataset. If you have an image classification task, for example, you must assign suitable class labels. You need to draw a bounding box around items to do an object detection task. Each input image pixel must be assigned the right class for a semantic segmentation task. This method necessitates manual labor, and labeling the training data can be quite costly. To appropriately categorize medical photos, for example, you'll need to hire pricey domain experts. Secondly, even gathering training photos can be difficult at times. Working with healthcare data is subject to a slew of legal limitations, and accessing it takes time and effort. Obtaining the training photos is sometimes possible, but it will be quite expensive. To obtain satellite images, for example, you must pay a satellite operator to capture the photos. You'll need an operator to drive a car and collect the necessary data to get photographs for road scene recognition.

The benchmark dataset in (Stoean et al., 2016) is considered during the experiments because the suggested E-CNN attempts to help in diagnosing colon cancer based on the histological pictures. The dataset was split into 20% for testing and 80% for training. All the suggested transfer learning models in E-CNN had the same value for Hyperparameters. The training and testing images were resized to 224 × 224 for comfort with the proposed transform learning models. The batch size was chosen as 16; the minimum learning rate was specified as min lr=0.000001. The learning rate was determined to be small enough to slow down learning in the models (Popa, 2021) (Kaur et al., 2021). The number of epochs was selected as 10. These models were trained by stochastic gradient descent (SGD) with momentum. All the proposed TL models employed cross entropy (CE)as the loss function. The cross-entropy is mainly utilized to estimate the distance between the prediction likelihood vector(E) and the one-hot-encoded ground truth label(T) (Boumaraf et al., 2021) probability vector (The following equation depicts the CE Eq. (2):

$$CE(E,T) = -\sum_{t=1} T_i \log E_i \tag{2}$$

where CE is used to determine how closely the output E corresponds to the baseline T. All of the proposed TL models also have a dropout layer to prevent over-fitting during training. In order to prevent units from overco-adapting, it randomly reduces activation throughout the training phase (Boumaraf et al., 2021). As is common when incorporating the dropout in deep learning models, the dropout parameter in this study was set to 0.3 to randomly remove the units.

4.3. Evaluation Criteria

Calculating the evaluation criteria will be based on using the confusion matrix; that contains information about the model's results, when applied to the testing dataset. With respect to the amount of data of the actual and predicted classes, Figure 3 shows an example of confusion matrix; Where the diagonal (TP and TN) refers to the set of points that were predicted correctly, and the diagonal (FN and FP) refers to the set of points that were predicted wrongly (Johansson et al., 2004)

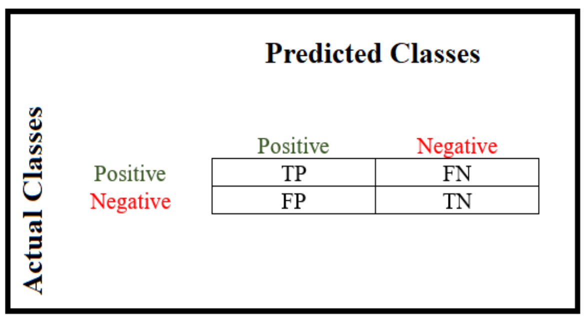


Fig. 3. Confusion Matrix

- The average classification accuracy: The correctly categorized TP and TN numbers combined with the criterion parameter, are generally referred to as accuracy. A technique's classification accuracy is measured in Eq. (3) as follows:

$$Acc = \frac{1}{M}\sum_{j=1}^{M} \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where M is the number of independent runs of the proposed ECNN with its individual.

- Average sensitivity: Sensitivity is also called recall. It represents the proportion of positive samples, which are efficiently determined as described in Eq. (4):

$$Sensitivity = \frac{1}{M}\sum_{j=1}^{M} \frac{TP}{TP+FN} * 100\% \tag{4}$$

The sensitivity ranges from [0, 1] on a scale. Zero displays the poorest possible categorization, whereas one displays the best classification. In order to get the necessary percentage, the sensitivity is multiplied by 100.

- Average Specificity: In a classification strategy, specificity serves as an evaluation indicator for negative samples. It specifically aims to quantify the proportion of the effectively categorized negative samples. Eq. (5) is used to compute specificity.

$$Specificity = \frac{1}{M}\sum_{j=1}^{M}\frac{TN}{TN+FP} * 100\% \quad (5)$$

5. Results and discussion

The proposed tooth caries ensemble deep models (TCE) and its people are presented in this subsection along with the experimental findings. To demonstrate the impact of the number of epochs on the individual and ensemble results, the individual pretrained models were trained using different numbers of epochs in this study (i.e., 10, 20, 30, 40, and 50). The effectiveness of each model (pretrained models) and the TCE was assessed using the accuracy, sensitivity, and specificity metrics over 10 runs. The suggested TCE was then contrasted with the most recent CNN models for classifying tooth caries. Tables 1 to 5 show the categorization performances of the TCE and its members using different numbers of epochs, for epochs 10, 20, 30, 40, and 50, respectively.

The average accuracy values of VGG16 model showed the lowest performance of 82% as shown in Table 1. However, this value increased dramatically when the number of epochs increased. for example, when number of epochs set to 20 (as illustrated in Table 2), the accuracy value of the VGG16 increased around 8%. while when the number of epochs set to 50 (Table 5), the accuracy increased to 94%. This means that, VGG16 model was trained very well and avoided the overfitting issue. Also, the figures of accuracy with epochs number (Figures 5, 9, 13, 17 and 21) prove this analysis. The classification performance of the VGG19 model has achieved around 89% when the number of epochs is equal to 10 as shown in Table 1. Fig. 6 shows that the accuracy value of the test data is not stable. This does not motivate us to increase the number of epochs to 30, where the VGG19 achieved an accuracy of around 93%, where the accuracy raises by around 4% compared to the same model when the number of epochs is equal to 10. These results show that when the number of epochs is more than what is required, the model is unable to perform adequately on a test dataset (Noon et al., 2020). For example, when number of epochs is equal to 50, more than enough, the accuracy was decreased around 2% when compared with the number of epochs 30 (Table 5).

The DenseNet121 model obtained around 86% accuracy when the number of epochs was 10. On the other hand, when the number of epochs reached 20, the accuracy increased by about 3%. However, the accuracy remained the same when the number of epochs reached 50. This indicates that Densenet121 reaches a stable state when the number of epochs is equal to 30 for the dentist domain. The last individual was the Xception model, which achieved the best accuracy among all other classifiers. When the number of epochs equals 10, Xception model obtained an accuracy value of around 94%, which is better than all other individuals. When the number of epochs increased to 50, the Xception model achieved a 96% accuracy value. Thus, the Xception model is considered the best individual in the proposed system. Accordingly, the results show that the Xception model is more suitable for dentist study than VGG16, VGG19, and DenseNet121.

The dentist's diagnosis performance can be improved by combining the basic (individual) classifiers into one model. In this study, the basis classifiers' judgments are combined using the majority voting and sum rule fusion approaches from ensemble learning. The test sample will be assigned to the class that has the highest sum after adding the probabilities from the four base classifiers. The results of the sum rule using various epochs are illustrated in Tables 1, 2, 3,4, and 5. From these tables, we can observe that the average accuracy for the Xception model increases from 93.9% to 96.5% when the number of epochs varied from 10 to 30. However, when the number of epochs reaches 40, the accuracy decreases to 95.2%, and 96.1% at 50 epochs. Looking at the sum rule, we notice that accuracy was highest with 97.2% at epochs 30. This indicates that the suitable number of epochs for an ensemble with a sum rule is 30.

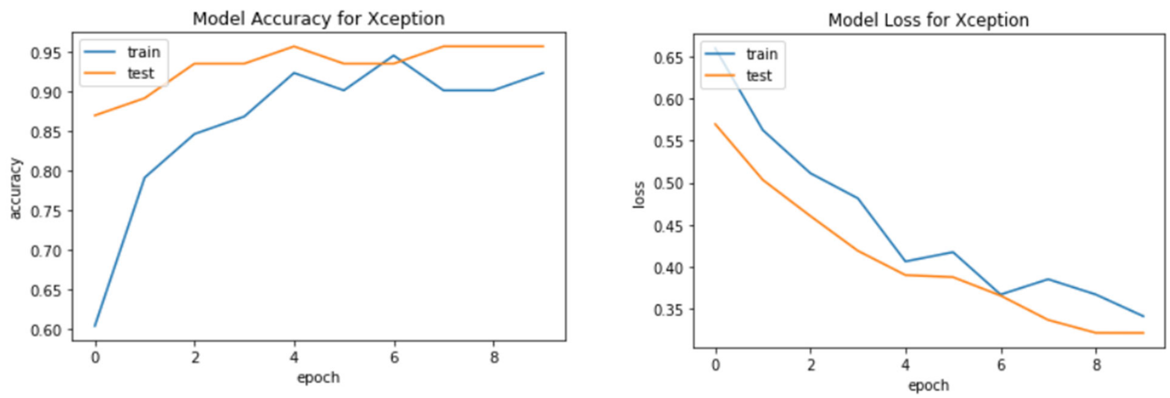
Overall, the experimental analysis of this research shows that the pretrained CNN can assist in the detection of dental caries from photography. Among the pretrained models, the Xception model shows good results using various performance measures: accuracy, sensitivity, and specificity. Moreover, the suggested pretrained ensemble methods perform better in terms of classification accuracy, sensitivity, and specificity thanks to the ensemble learning techniques (sum rule and majority voting). The suggested framework for ensemble classification in dental caries utilizing images is straightforward but effective.

5. Conclusion and Future Work

This paper proposed two ensemble learning models based on (Ensemble (majority voting) and Ensemble (sum)) for detecting the caries lesions in photographs. The proposed ensemble models utilized four different deep CNN learning models (Xception, VGG16, VGG19, and DenseNet121), to extract the essential features from the input images. The proposed ensemble models were approved over the dental benchmark dataset. The obtained results showed the superiority of the proposed models; actually, in all recorded results, the Ensemble (Sum) values were optimal. This indicates that these models can be extended in future works for another medical domain.

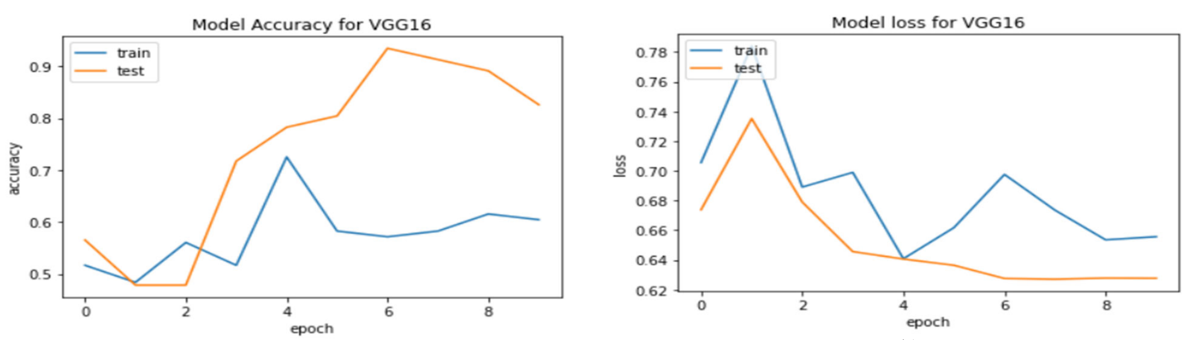
Table 1
Accuracy, sensitivity, and specificity of the proposed modified pretrained CNN models for teeth caries detection for 10 Epochs

Model used	Accuracy		Sensitivity		Specificity	
	AVG	STD	AVG	STD	AVG	**STD**
Xception	0.939	3.57	0.941	3.395	0.995	**1.364**
VGG16	0.828	0.135	0.829	0.128	0.855	**0.145**
VGG19	0.889	0.088	0.890	0.086	0.914	**0.141**
DenseNet121	0.863	0.07	0.868	0.067	0.973	**0.058**
Ensemble (Sum)	0.957	0.014	0.958	0.014	0.986	**0.021**
Ensemble (Majority voting)	**0.924**	**2.957**	**0.927**	**2.841**	**0.995**	**1.364**



a) b)

Fig. 4. The suggested improved pretrained CNN Xception model's learning curves for (a) training and testing accuracy and (b) training and testing loss for detecting tooth decay over 10 epochs.

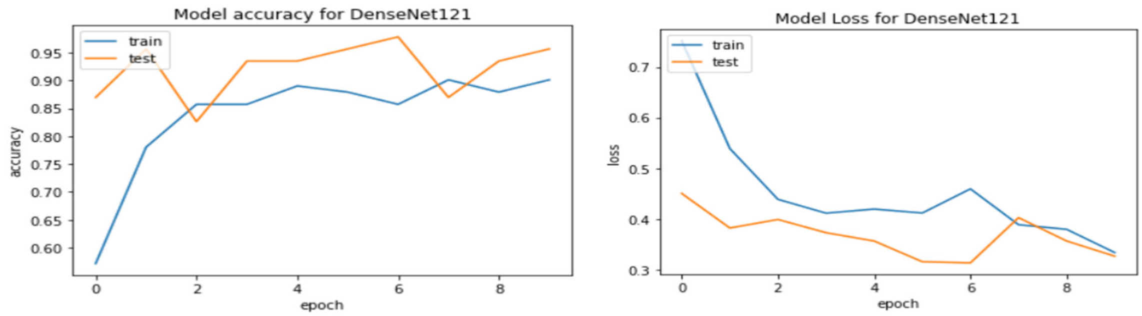


a) b)

Fig. 5. Learning curves for the proposed modified pretrained CNN VGG16 model for teeth caries detection for 10 Epochs for (a) training and testing accuracy and (b) training and testing loss.

a) b)

Fig. 6. Learning curves for the proposed modified pretrained CNN VGG19 model for detecting tooth decay over 10 epochs for (a) training and testing accuracy and (b) training and testing loss.

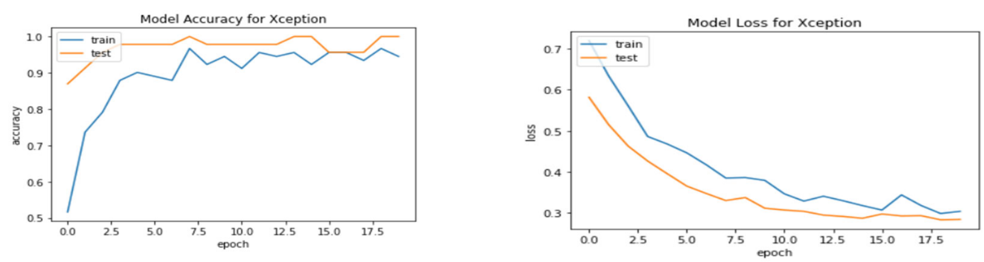


a) b)

Fig. 7. Learning curves for the proposed modified pretrained CNN DenseNet model for detecting tooth decay across 10 epochs for (a) training and testing accuracy and (b) training and testing loss.

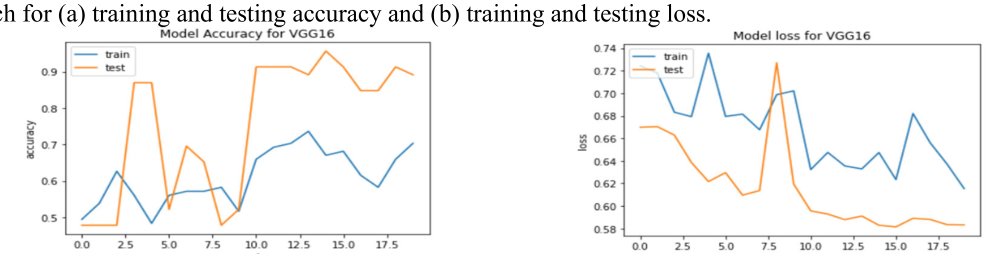
Table 2
Accuracy, sensitivity and specificity of the proposed modified pretrained CNN models for teeth caries detection for 20 Epochs

Model used	Accuracy		Sensitivity		Specificity	
	AVG	STD	AVG	STD	AVG	STD
Xception	0.948	2.421	0.95	2.32	1.000	0.000
VGG16	0.893	0.063	0.897	0.061	0.968	0.029
VGG19	0.852	0.109	0.858	0.104	0.986	0.021
DenseNet121	0.896	0.044	0.900	0.042	0.982	0.030
Ensemble (Sum)	0.963	0.014	0.964	0.013	0.995	0.014
Ensemble (Majority voting)	0.902	4.038	0.906	3.87	1.000	0.000



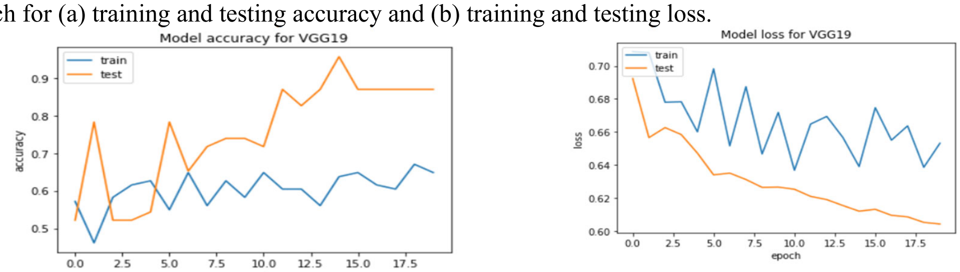
a) b)

Fig. 8. Learning curves for the proposed modified pretrained CNN Xception model for teeth caries diagnosis for 20 Epoch for (a) training and testing accuracy and (b) training and testing loss.



a) b)

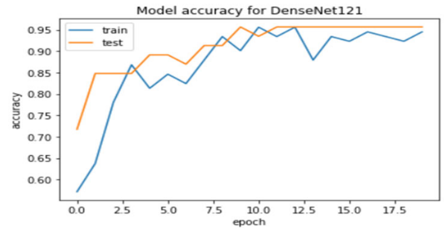
Fig. 9. Learning curves of the proposed modified pretrained CNN VGG16 model for teeth caries detection for 20 Epoch for (a) training and testing accuracy and (b) training and testing loss.



b)

a)

Fig. 10. Learning curves for the proposed modified pretrained CNN VGG19 model for tooth caries detection for 20 Epoch, including (a) training and testing accuracy and (b) training and testing loss.



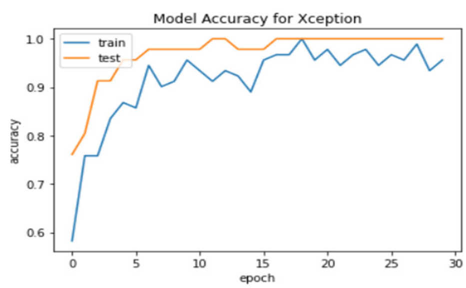
a)

b)

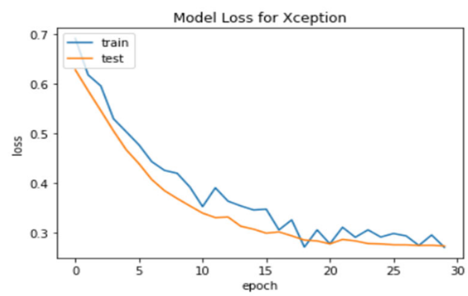
Fig. 11. Learning curves for the proposed modified pretrained CNN DenseNet model for detecting tooth decay over a 20-epoch period for (a) training and testing accuracy and (b) training and testing loss.

Table 3
Accuracy, sensitivity and specificity of the proposed modified pretrained CNN models for teeth caries detection for 30 Epochs

Model used	Accuracy		Sensitivity		Specificity	
	AVG	STD	AVG	STD	AVG	STD
Xception	0.965	1.065	0.966	1.038	0.995	1.364
VGG16	0.902	0.065	0.904	0.061	0.941	0.058
VGG19	0.937	0.025	0.938	0.024	0.964	0.027
DenseNet121	0.891	0.034	0.895	0.032	0.991	0.018
Ensemble (Sum)	0.972	0.010	0.973	0.010	0.991	0.018
Ensemble (Majority voting)	0.935	2.174	0.937	2.065	0.995	1.364

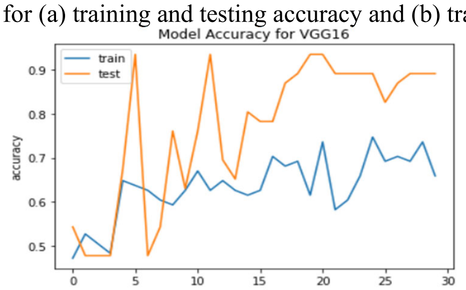


a)

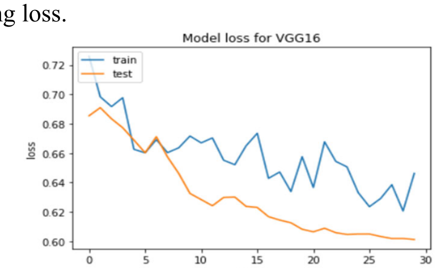


b)

Fig. 12. Learning curves for the proposed modified pretrained CNN Xception model for teeth caries detection for 30 Epoch for (a) training and testing accuracy and (b) training and testing loss.

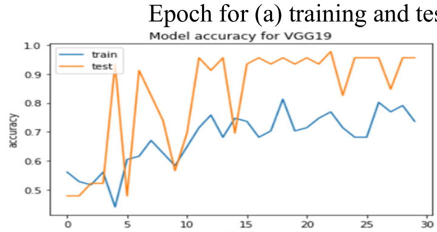


a)

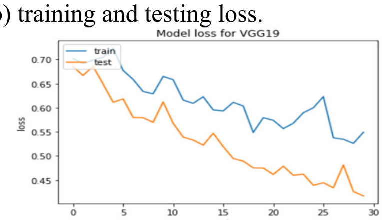


b)

Fig. 13. Learning curves for the proposed modified pretrained CNN VGG16 model for teeth caries detection for 30 Epoch for (a) training and testing accuracy and (b) training and testing loss.

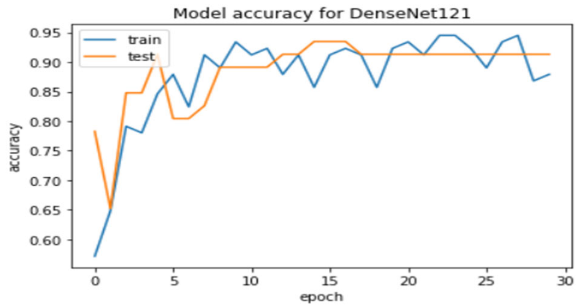


a)



b)

Fig. 14. Learning curves for the proposed modified pretrained CNN VGG19 model for teeth caries detection for 30 Epoch for (a) training and testing accuracy and (b) training and testing loss.



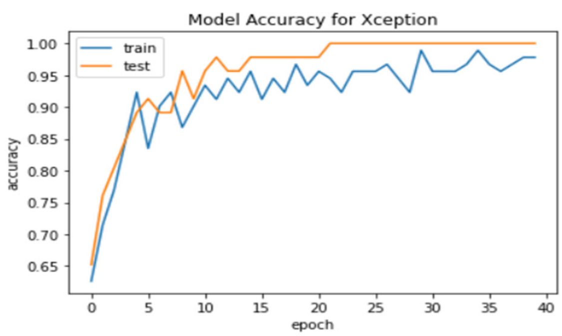
a)

b)

Fig. 15. Learning curves for the proposed modified pretrained CNN DenseNet model for tooth caries detection for 30 Epoch for (a) training and testing accuracy and (b) training and testing loss.

Table ٤
Accuracy, sensitivity and specificity of the proposed modified pretrained CNN models for teeth caries detection for 40 Epochs

Model used	Accuracy		Sensitivity		Specificity	
	AVG	STD	AVG	STD	AVG	STD
Xception	0.952	2.535	0.954	2.430	1.000	0.000
VGG16	0.937	0.064	0.938	0.061	0.959	0.014
VGG19	0.917	0.043	0.920	0.041	0.982	0.022
DenseNet121	0.893	0.042	0.897	0.041	0.977	0.042
Ensemble (Sum)	0.965	0.011	0.966	0.010	0.995	0.014
Ensemble (Majority voting)	0.928	2.758	0.931	2.669	0.995	1.364

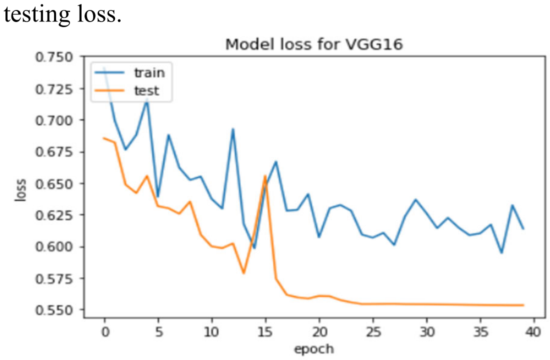


a)

b)

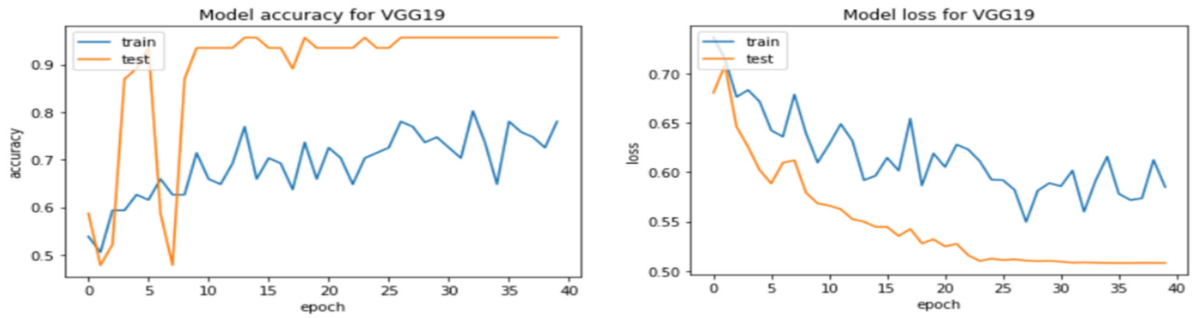
Fig. 16. Learning curves of the proposed modified pretrained CNN Xception model for tooth caries diagnosis for 40 Epoch for (a) training and testing accuracy and (b) training and testing loss.

a)



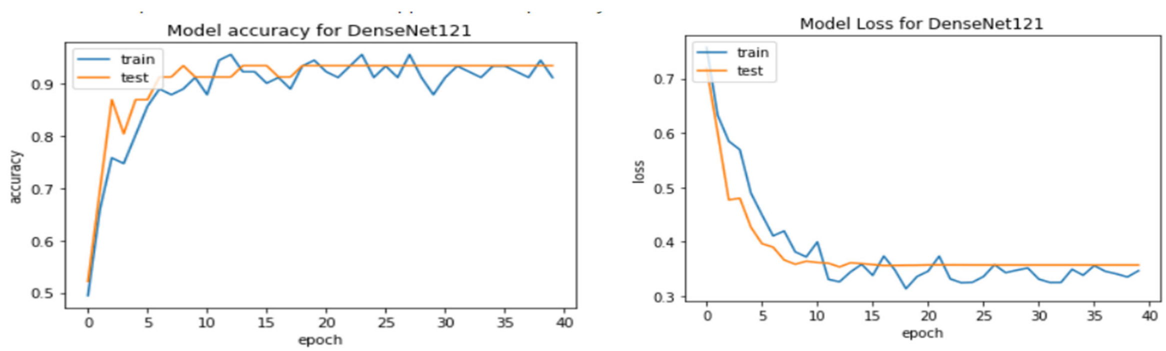
b)

Fig. 17. Learning curves for the proposed modified pretrained CNN VGG16 model for teeth caries detection over 40 epochs for (a) training and testing accuracy and (b) training and testing loss.



a) b)

Fig. 18. Learning curves of the proposed modified pretrained CNN VGG19 model for tooth caries detection for 40 Epoch for (a) training and testing accuracy and (b) training and testing loss.

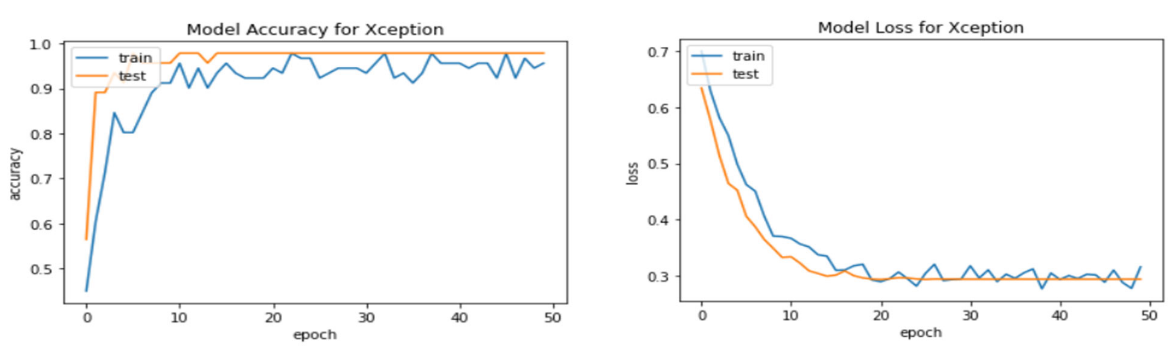


a) b)

Fig. 19. Learning curves for the proposed modified pretrained CNN DenseNet model for tooth caries detection for 40 Epoch for (a) training and testing accuracy and (b) training and testing loss.

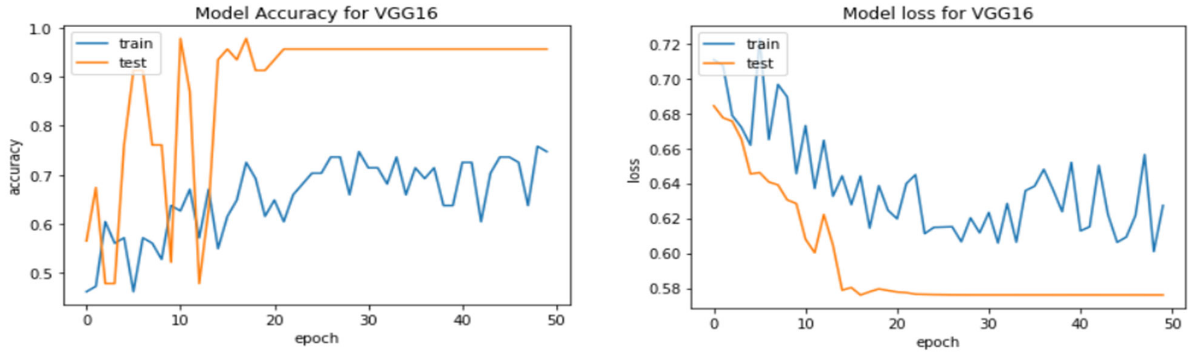
Table 5
Accuracy, sensitivity and specificity of the proposed modified pretrained CNN for teeth caries detection for 50 Epochs

Model used	Accuracy		Sensitivity		Specificity	
	AVG	STD	AVG	STD	AVG	STD
Xception	0.961	1.304	0.963	1.250	1.000	0.000
VGG16	0.943	0.049	0.944	0.047	0.955	0.020
VGG19	0.926	0.033	0.928	0.030	0.964	0.034
DenseNet121	0.874	0.037	0.878	0.037	0.973	0.042
Ensemble (Sum)	0.963	0.017	0.964	0.016	0.995	0.014
Ensemble (Majority voting)	0.935	2.572	0.937	2.465	1.000	0.000



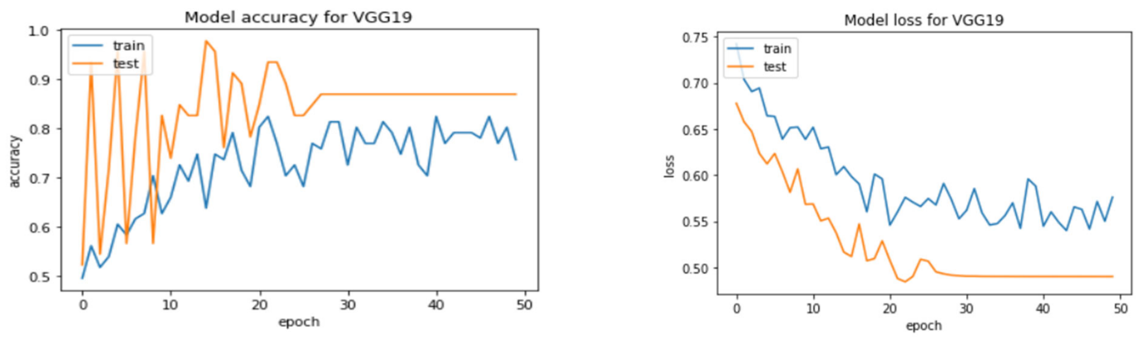
a) b)

Fig. 20. Learning curves for the proposed modified pretrained CNN Xception model for tooth caries diagnosis for 50 Epoch for (a) training and testing accuracy and (b) training and testing loss.



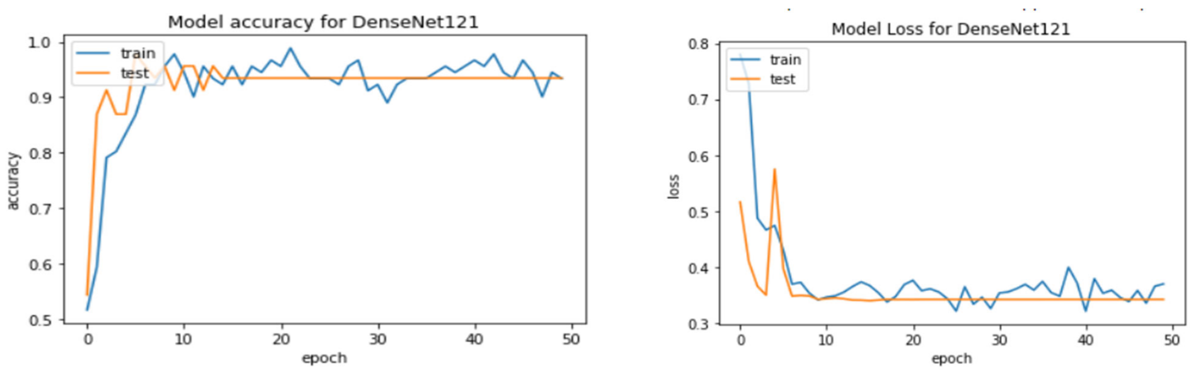
a) b)

Fig. 21. Learning curves for the proposed modified pretrained CNN VGG16 model for tooth caries detection for 50 Epoch for (a) training and testing accuracy and (b) training and testing loss.



a) b)

Fig. 22. Learning curves for the proposed modified pretrained CNN VGG19 model for tooth caries detection for 50 Epoch for (a) training and testing accuracy and (b) training and testing loss.



a) b)

Fig. 23. Learning curves for the proposed modified pretrained CNN DenseNet model for detecting tooth decay over a period of 50 epochs for (a) training and testing accuracy and (b) training and testing loss.

References

Boumaraf, S., Liu, X., Zheng, Z., Ma, X., & Ferkous, C. (2021). A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images. *Biomedical Signal Processing and Control*, *63*, 102192.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).

Choudhary, A., Raj, G., Agrawal, A. P., Sawhney, H., Nand, P., & Bhargava, D. (2021, December). An effective approach for classification of dental caries using Convolutional Neural Networks. In *2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART)* (pp. 204-209). IEEE.

Cunha, S. R., Maravic, T., Comba, A., Ramos, P. A., Tay, F. R., Pashley, D. H., ... & Breschi, L. (2020). In vivo and in vitro radiotherapy increased dentin enzymatic activity. *Journal of Dentistry*, *100*, 103429.

Haghanifar, A., Majdabadi, M. M., & Ko, S. B. (2020). Paxnet: Dental caries detection in panoramic x-ray using ensemble transfer learning and capsule classifier. *arXiv preprint arXiv:2012.13666*.

Haghanifar, A. (2022). *Automated Teeth Extraction and Dental Caries Detection in Panoramic X-ray* (Doctoral dissertation, University of Saskatchewan).

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

Imak, A., Celebi, A., Siddique, K., Turkoglu, M., Sengur, A., & Salam, I. (2022). Dental caries detection using score-based multi-input deep convolutional neural network. *IEEE Access*, *10*, 18320-18329.

Johansson, U., Niklasson, L., & König, R. (2004, June). Accuracy vs. comprehensibility in data mining models. In *Proceedings of the seventh international conference on information fusion* (Vol. 1, pp. 295-300). Stockholm, Sweden: Elsevier.

Kaur, T., & Gandhi, T. K. (2020). Deep convolutional neural networks with transfer learning for automated brain image classification. *Machine vision and applications*, *31*(3), 20.

Lian, L., Zhu, T., Zhu, F., & Zhu, H. (2021). Deep learning for caries detection and classification. *Diagnostics*, *11*(9), 1672.

Manna, A., Kundu, R., Kaplun, D., Sinitca, A., & Sarkar, R. (2021). A fuzzy rank-based ensemble of CNN models for classification of cervical cytology. *Scientific Reports*, *11*(1), 14538.

Noon, S. K., Amjad, M., Qureshi, M. A., & Mannan, A. (2020, November). Overfitting mitigation analysis in deep learning models for plant leaf disease recognition. In *2020 IEEE 23rd International Multitopic Conference (INMIC)* (pp. 1-5). IEEE.

Popa, L. (2021). A statistical framework for evaluating convolutional neural networks. Application to colon cancer. *Annals of the University of Craiova-Mathematics and Computer Science Series*, *48*(1), 159-166.

Revilla-León, M., Young, K., Sicilia, E., Cho, S. H., & Kois, J. C. (2022). Influence of definitive and interim restorative materials and surface finishing on the scanning accuracy of an intraoral scanner. *Journal of dentistry*, *120*, 104114.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sinha, A., Ramish, M., Kumari, S., Jha, P., & Tiwari, M. K. (2022, January). ANN-ANT-LION-MLP Ensemble Transfer Learning Based Classifier for Detection and Classification of Oral Disease Severity. In *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 530-535). IEEE.

Sonavane, A., Yadav, R., & Khamparia, A. (2021). Dental cavity classification of using convolutional neural network. In *IOP conference series: materials science and engineering* (Vol. 1022, No. 1, p. 012116). IOP Publishing.

Stark, B., & Samarah, M. (2019, December). Ensemble and Deep Learning for Real-time Sensors Evaluation of algorithms for real-time sensors with application for detecting brushing location. In *2019 IEEE 5th International Conference on Computer and Communications (ICCC)* (pp. 555-559). IEEE.

Stoean, C., Stoean, R., Sandita, A., Ciobanu, D., Mesina, C., & Gruia, C. L. (2016). Svm-based cancer grading from histopathological images using morphological and topological features of glands and nuclei. In *Intelligent interactive multimedia systems and services 2016* (pp. 145-155). Springer International Publishing.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

Xue, D., Zhou, X., Li, C., Yao, Y., Rahaman, M. M., Zhang, J., ... & Sun, H. (2020). An application of transfer learning and ensemble learning techniques for cervical histopathology image classification. *IEEE Access*, *8*, 104603-104618.

Zhu, H., Cao, Z., Lian, L., Ye, G., Gao, H., & Wu, J. (2022). CariesNet: a deep learning approach for segmentation of multi-stage caries lesion from oral panoramic X-ray image. *Neural Computing and Applications*, 1-9.