

Toward development of automated grading system for carious lesions classification using deep learning and OCT imaging

Hassan S. Salehi^{a,*}, Majd Barchini^a, Qingguang Chen^{a,b}, Mina Mahdian^c

^aDepartment of Electrical and Computer Engineering, California State University, Chico, CA, USA

^bSchool of Automation, Hangzhou Dianzi University, Hangzhou, Zhejiang, China

^cSchool of Dental Medicine, Stony Brook University, Stony Brook, NY, USA

[*hsalehi@csuchico.edu](mailto:hsalehi@csuchico.edu)

ABSTRACT

Dental caries remains the most prevalent chronic disease in both children and adults. Optical coherence tomography (OCT) is a noninvasive optical imaging modality extensively utilized to image oral samples to diagnose carious lesions, but detecting early stage dental caries with high-level accuracy remains challenging. Deep learning models have been employed to classify OCT images for various healthcare applications. In this paper, human tooth specimens were imaged ex vivo using OCT imaging systems, and a three-class grading system based on deep learning model for detection and classification of carious lesions was developed. Human extracted premolar and molar teeth were collected and categorized into three classes, Grade 0: healthy (non-carious teeth), Grade 1: early-stage caries (caries extending into enamel), and Grade 2: late-stage caries (caries extending into dentin). For OCT imaging, a spectral-domain OCT system and a swept-source OCT system were utilized. Advanced image processing and augmentation techniques were performed to prepare the image data and generate additional examples of each class prior to the deep learning process. For deep learning, ten deep convolutional neural networks (CNN) architectures were investigated to determine the optimal numbers of convolutional and fully connected layers for the classification tasks. The diagnostic accuracy, sensitivity, specificity, positive predictive value, and negative predictive value were calculated for detection and diagnostic performances of the CNN models. This study is a step forward in the development of automated deep learning/OCT imaging system for early dental caries diagnosis.

Keywords: optical coherence tomography, image processing, deep learning, convolutional neural networks, classification, carious lesions.

1. INTRODUCTION

Detection of carious lesions at the initial stages of demineralization, can result in the implementation of non-surgical preventive approaches to reverse the demineralization process. Therefore, a reliable diagnostic imaging modality that can effectively identify and quantify the extent of caries in early stages with high sensitivity and specificity and minimal invasiveness, is essential ^{1,2}. The conventional approach for diagnosing dental caries is clinical examination supplemented by radiographs. However, studies based on the clinical and radiographic examination methods often show low sensitivity and high specificity. Moreover, by the time a lesion is visualized in clinical or radiographic examination, it is believed to have advanced to the extent that non-invasive preventive measures may no longer lead to remineralization of the lesion ^{2,3}. Optical coherence tomography (OCT) is a noninvasive imaging modality based on low-coherence interferometry that utilizes non-ionizing near-infrared laser to provide micrometer-resolution images (1–10 μm), approximately 100 times higher than conventional ultrasound imaging systems ⁴⁻¹². In deep learning ¹³⁻¹⁶, a convolutional neural network (CNN) is the most commonly applied to analyzing biomedical and medical imaging data. Recent studies have demonstrated the CNN application for complex medical image analysis ¹⁷⁻²¹.

In our previous publication, we presented a novel approach combining OCT imaging modality and deep learning CNN model for the detection of occlusal carious lesions (binary classification: non-carious or carious teeth). To the best of our knowledge, that study was the first one reporting deep learning-based classification of ex vivo OCT images of human carious and non-carious lesions for detection of dental caries ²²⁻²⁴.

In this study, a three-class grading system of OCT images based on deep learning model for detection and classification of carious lesions was developed. Human extracted premolar and molar teeth were collected and categorized into three classes, Grade 0: healthy (non-carious teeth), Grade 1: early-stage caries (caries extending into enamel), and Grade 2: late-stage caries (caries extending into dentin). For OCT imaging, a spectral-domain OCT system and a swept-source OCT system were utilized. Advanced image processing and augmentation techniques were performed to prepare the OCT image data and generate additional examples of each class prior to the deep learning process. For deep learning, ten deep convolutional neural networks architectures were investigated to determine the optimal numbers of convolutional and fully connected layers for the classification tasks.

2. METHODS

2.1 Overall automated grading system

Figure 1 demonstrates the overall automated grading system based on OCT/deep learning for early dental caries diagnosis. Human tooth specimens were imaged ex vivo using OCT imaging systems, and a three-class grading system based on deep learning model for detection and classification of carious lesions was developed.

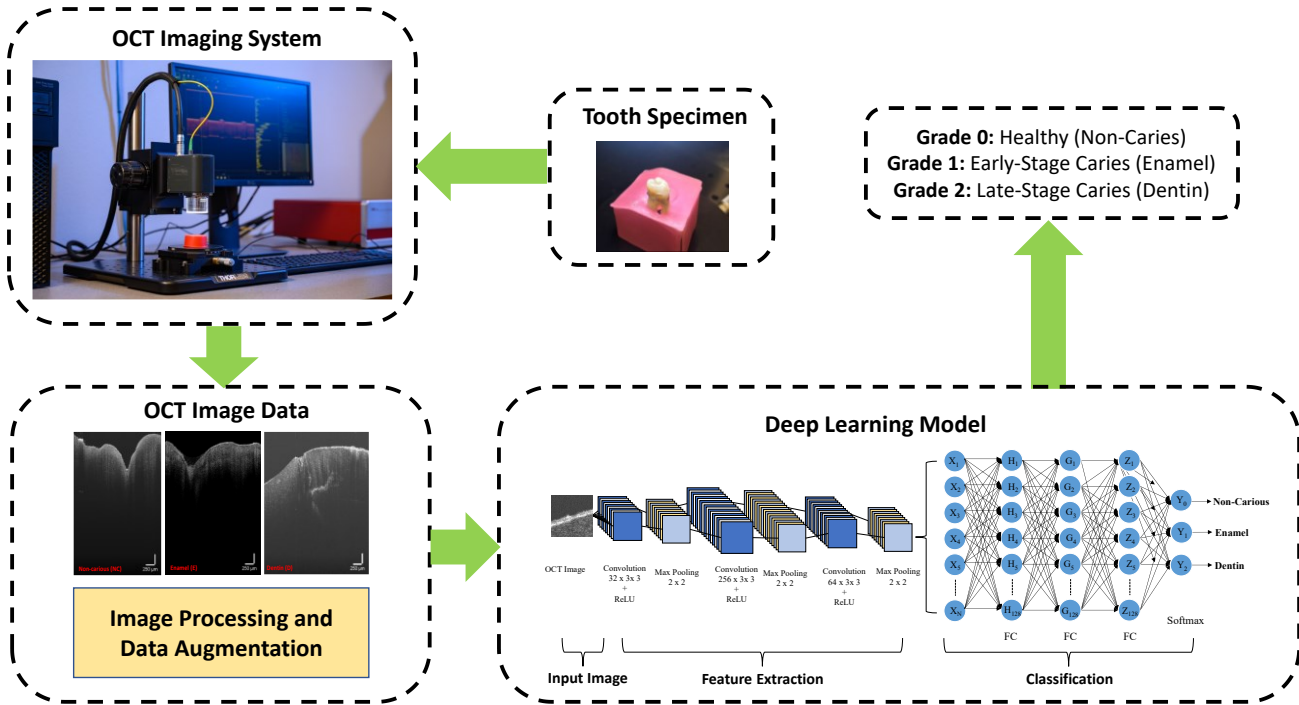


Figure 1 OCT/deep learning-based grading system for early dental caries diagnosis.

2.2 Human tooth specimens

Institutional review board (IRB) approvals were obtained from SUNY Stony Brook University and California State University, Chico to collect a total of 81 human extracted premolar and molar teeth and categorized into three classes, Grade 0: healthy (non-carious teeth), Grade 1: early-stage caries (caries extending into enamel), and Grade 2: late-stage caries (caries extending into dentin). The teeth were disinfected and stored in distilled water.

2.3 Ex vivo OCT imaging and data acquisition

For OCT imaging, two different systems were utilized; spectral-domain OCT system from the TELESTO-series (Thorlabs Inc., Newton, NJ, USA) operating at 1300 nm center wavelength with A-scan rate of 5.5-76 kHz, imaging depth of 3.5 mm, and axial resolution of 5.5 μm , and swept-source OCT system from the VEGA-series (Thorlabs Inc., Newton, NJ, USA) operating at 1300 nm center wavelength with A-scan rate of 200 kHz, imaging depth of 8 mm, and axial resolution of 16 μm . The surfaces of the teeth were kept hydrated for optimal light penetration and refraction. To acquire images with minimum inhomogeneity, imaging was performed multiple times at different points. OCT images with the least heterogeneous presentation were imported and saved in TIFF format. Figure 2 shows examples for OCT images of three classes and photos of the corresponding specimens.

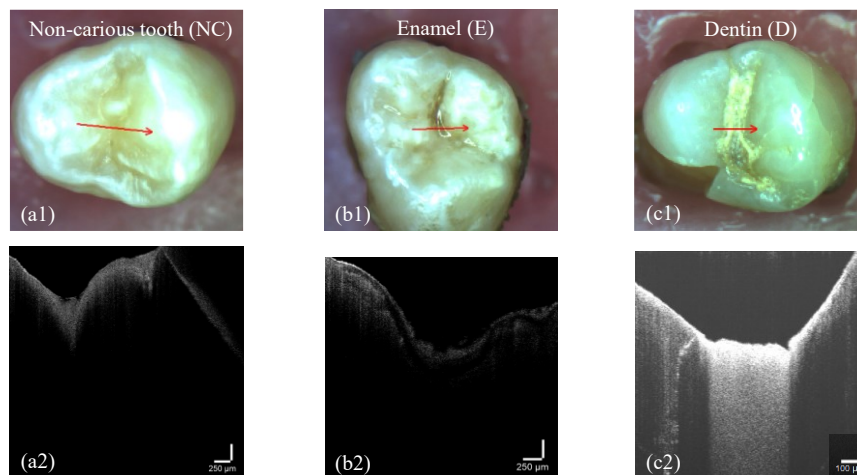


Figure 2 (a1-c1) Tooth specimens; (a2-c2) corresponding OCT images.

2.4 Image processing and Augmentation

Image processing techniques were performed to obtain a database with consistent image format and one fixed size. Deep neural networks require a large amount of training data to effectively learn, where collection of such training data is often expensive and laborious. Data augmentation overcomes this issue by artificially inflating the training set with label preserving transformations. We have applied image augmentation by perturbing an image using transformations that leave the underlying class unchanged (e.g. cropping and flipping) in order to generate additional examples of the class. Image augmentation can be applied at training time, at test time, or both. The augmented samples can either be taken as-is or combined to form a single feature, e.g. using sum/max-pooling or stacking. In our processing, the rotational angles were in the range of 0° - 345° with 15° increase, and the images were flipped horizontally. Then, the images were cropped to the size of 90×90 . In Fig. 3, samples of OCT image augmentation are illustrated.

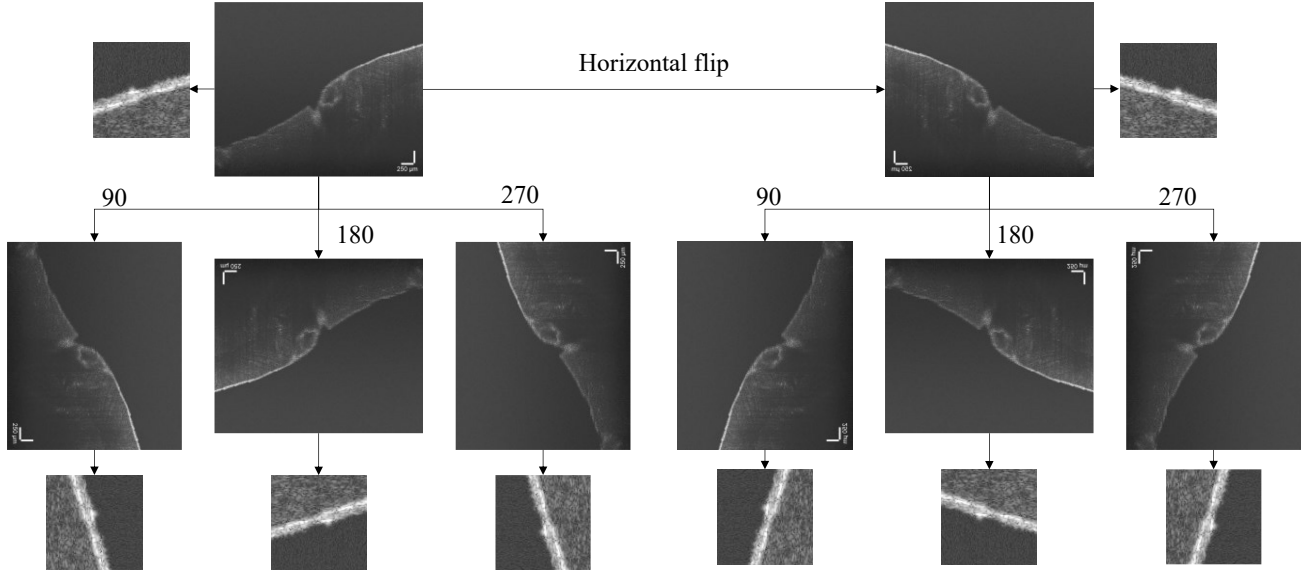


Figure 3 OCT image data augmentation.

2.5 Deep learning models

The deep learning employs computational models which are composed of a series of transforming and processing layers to learn representations of data with multiple levels of abstraction. Ten CNN architectures were investigated to determine the optimal numbers of convolutional and fully connected layers for OCT image classification tasks. The deep learning models were implemented using TensorFlow opensource library. The training set was split into mini-batches, with 10 images per batch. Given a batch of training patches, the CNN uses convolution and pooling layers to extract features and then classify each patch based on the probabilities from the softmax classification layer. After that, the CNN calculates the error between the classification result and the reference label, and then utilizes the backpropagation process to tune all the layer weights to minimize this error using Nadam optimizer with 0.001 learning rate²⁴. The above process will be repeated several epochs, until the whole CNN model becomes convergent. In this research, 75% of the imaging data was utilized for training and 25% for testing.

3. RESULTS

To investigate the CNN structures, we performed two studies. In the first study, the CNN models had two fully connected (FC) layers and three convolutional layers with five combinations obtained by changing the number of filters in the last convolutional layer of the feature extraction part of the CNN model, as demonstrated in Fig. 4 and Table 1. The first convolutional layer (Conv1) has 32 filters of size 3x3 with Rectified Liner Unite (ReLU) layer followed by a max-pooling layer of size 2x2. The second convolutional layer (Conv2) with 256 filters of size 3x3 with ReLU layer followed by a max-pooling layer of size 2x2. In the third convolutional layer (Conv3), we investigated different numbers of filters (N in Fig. 4) of size 3x3 with ReLU layer and followed by a max-pooling layer of size 2x2. After feature extraction, the final feature map was flattened to a 1D vector, connected to the first Fully Connected (FC) layer with 128 neurons and ReLU activation function followed by drop out layer. The first FC layer's outputs are connected to the second FC layer with 128 neurons and ReLU activation function followed by drop out layer. The classification layer has three neurons with a softmax activation function that predicts the input image grade (0: healthy, 1: enamel, 2: dentin). All neurons weights in the network were randomly initialized from a Gaussian distribution with a 0 mean and a standard deviation of 0.01.

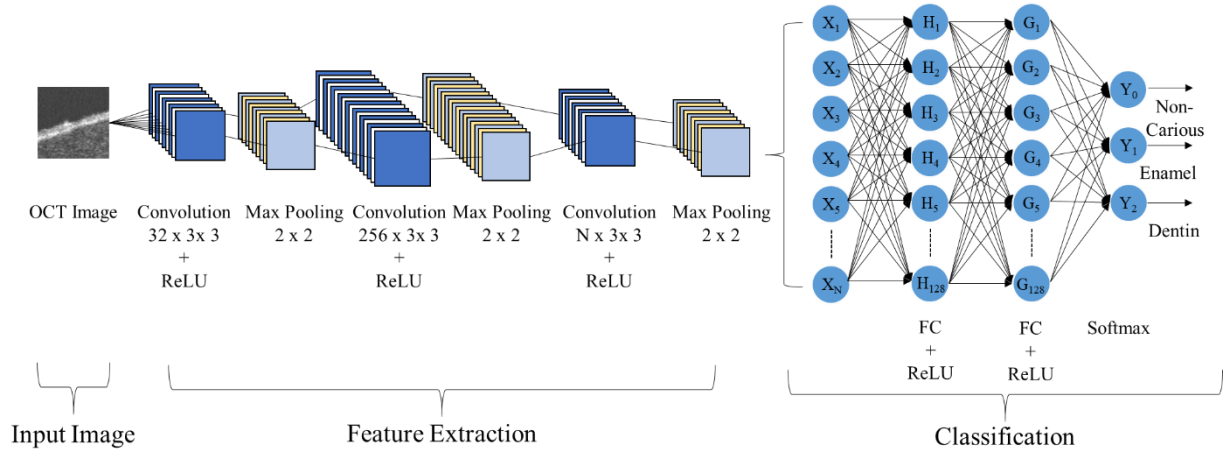


Figure 4 The CNN architecture for the first study.

For quantitative analysis of the experimental results, several performance metrics were considered, including diagnostic accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). To do this, we also used the variables true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Based on the results summarized in Table 1, the best CNN architecture in the first study had 3 convolutional layers with 32, 256, 128 filters and 2 fully connected layers with 128 neurons in each layer.

Table 1 Classification results of *ex vivo* OCT images using 5 CNN models

conv1	conv2	conv3	FC1	FC2	FC3	OL	Categories	Training					Testing				
								Accuracy	Sensitivity	Specificity	PPV	NPV	Accuracy	Sensitivity	Specificity	PPV	NPV
32	256	-	128	128	-	Softmax	Non Carious	95.90%	92.31%	98.02%	96.50%	95.58%	89.36%	82.32%	93.54%	88.31%	89.92%
							Enamel	96.02%	96.38%	95.79%	93.79%	97.57%	89.35%	90.26%	88.75%	84.14%	93.23%
							Dentin	97.77%	96.28%	98.22%	94.20%	98.87%	93.90%	85.91%	96.28%	87.32%	95.82%
32	256	32	128	128	-	Softmax	Non Carious	82.10%	94.00%	75.09%	68.99%	95.50%	81.06%	93.03%	73.96%	67.93%	94.71%
							Enamel	83.37%	78.27%	86.73%	79.58%	85.80%	82.52%	77.04%	86.13%	78.61%	85.01%
							Dentin	84.26%	38.32%	98.08%	85.75%	84.09%	84.06%	37.08%	98.07%	85.11%	83.94%
32	256	64	128	128	-	Softmax	Non Carious	92.90%	84.57%	97.81%	95.80%	91.49%	91.01%	81.38%	96.72%	93.64%	89.76%
							Enamel	94.07%	92.63%	95.03%	92.48%	95.12%	92.00%	89.90%	93.38%	89.99%	93.33%
							Dentin	94.34%	97.03%	93.53%	81.86%	99.05%	93.11%	95.69%	92.34%	78.84%	98.63%
32	256	128	128	128	-	Softmax	Non Carious	92.85%	85.45%	97.22%	94.77%	91.89%	90.88%	81.86%	96.23%	92.80%	89.95%
							Enamel	94.29%	96.23%	93.01%	90.09%	97.39%	92.41%	94.59%	90.98%	87.39%	96.22%
							Dentin	96.39%	94.23%	97.04%	90.54%	98.24%	95.16%	91.88%	96.14%	87.65%	97.54%
32	256	256	128	128	-	Softmax	Non Carious	81.34%	64.96%	91.00%	80.98%	81.50%	80.27%	64.45%	89.65%	78.70%	80.97%
							Enamel	83.82%	60.33%	99.33%	98.35%	79.13%	82.72%	58.16%	98.96%	97.36%	78.15%
							Dentin	76.28%	97.82%	69.80%	49.35%	99.07%	75.91%	97.11%	69.59%	48.77%	98.78%

In the second study, the CNN models had three fully connected (FC) layers and three convolutional layers with five combinations obtained by changing the number of filters in the last convolutional layer of the feature extraction part of the CNN model, as demonstrated in Fig. 5 and Table 2. The first convolutional layer (Conv1) has 32 filters of size 3x3 with Rectified Liner Unite (ReLU) layer followed by a max-pooling layer of size 2x2. The second convolutional layer (Conv2) with 256 filters of size 3x3 with ReLU layer followed by a max-pooling layer of size 2x2. In the third convolutional layer (Conv3) similar to the first study, we investigated different numbers of filters (N in Fig. 5) of size 3x3 with ReLU layer and followed by a max-pooling layer of size 2x2. After feature extraction, the final feature map was flattened to a 1D vector, connected to the first Fully Connected (FC) layer with 128 neurons and ReLU activation function followed by drop out layer. The first FC layer's outputs are connected to the second FC layer with 128 neurons

and ReLU activation function followed by drop out layer. The classification layer has three neurons with a softmax activation function that predicts the input image grade (0: healthy, 1: enamel, 2: dentin).

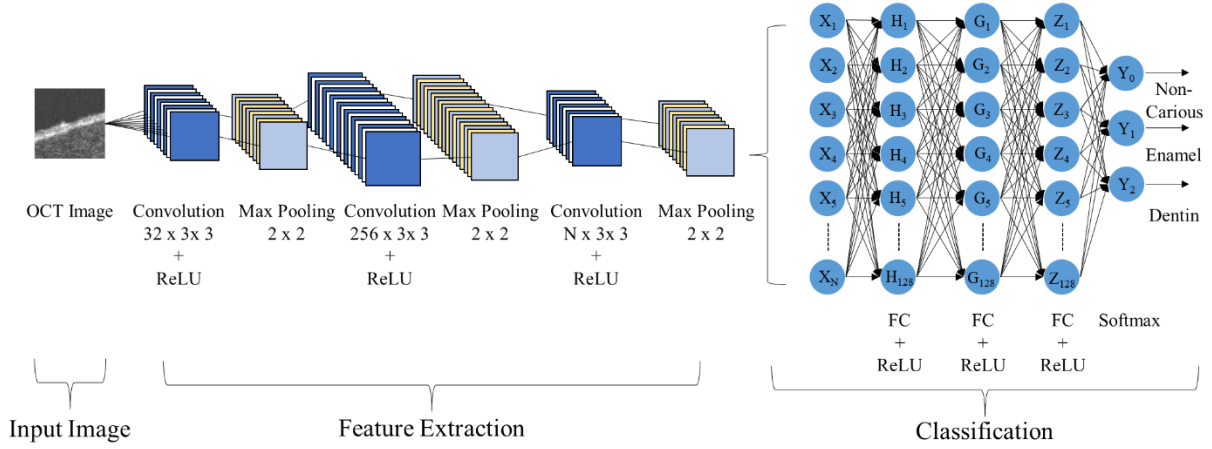


Figure 5 The CNN architecture for the second study.

Again, for quantitative analysis of the experimental results, the same performance metrics were considered, namely diagnostic accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). To do this, we used the variables true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

Based on the results summarized in Table 2, the best CNN architecture in the second study had 3 convolutional layers with 32, 256, 64 filters and 3 fully connected layers with 128 neurons.

Table 2 Classification results of *ex vivo* OCT images using ten CNN models

conv1	conv2	conv3	FC1	FC2	FC3	OL	Categories	Training					Testing				
								Accuracy	Sensitivity	Specificity	PPV	NPV	Accuracy	Sensitivity	Specificity	PPV	NPV
32	256	-	128	128	128	Softmax	Non Carious	93.95%	89.84%	96.37%	93.59%	94.14%	87.89%	81.80%	91.50%	85.09%	89.45%
							Enamel	92.85%	96.61%	90.36%	86.88%	97.58%	86.42%	90.94%	83.42%	78.39%	93.30%
							Dentin	95.55%	83.95%	99.04%	96.32%	95.35%	92.24%	72.36%	98.16%	92.15%	92.26%
32	256	32	128	128	128	Softmax	Non Carious	92.51%	89.77%	94.12%	90.00%	93.98%	91.02%	87.95%	92.84%	87.92%	92.85%
							Enamel	92.26%	91.38%	92.84%	89.40%	94.22%	90.69%	89.73%	91.33%	87.25%	93.08%
							Dentin	96.00%	89.65%	97.91%	92.82%	96.92%	95.04%	86.71%	97.52%	91.24%	96.10%
32	256	64	128	128	128	Softmax	Non Carious	93.60%	97.07%	91.55%	87.14%	98.15%	92.02%	95.95%	89.69%	84.66%	97.39%
							Enamel	94.42%	89.15%	97.91%	96.57%	93.18%	92.91%	86.94%	96.86%	94.83%	91.82%
							Dentin	97.54%	92.16%	99.16%	97.07%	97.68%	96.58%	88.94%	98.85%	95.85%	96.77%
32	256	128	128	128	128	Softmax	Non Carious	91.94%	90.27%	92.93%	88.28%	94.18%	90.43%	88.39%	91.65%	86.25%	93.01%
							Enamel	92.70%	85.70%	97.32%	95.49%	91.15%	90.94%	82.98%	96.21%	93.54%	89.53%
							Dentin	95.16%	96.54%	94.75%	84.68%	98.91%	94.06%	94.85%	93.83%	82.09%	98.39%
32	256	256	128	128	128	Softmax	Non Carious	91.04%	87.99%	92.84%	87.87%	92.91%	89.31%	85.62%	91.50%	85.66%	91.48%
							Enamel	92.56%	83.78%	98.35%	97.11%	90.18%	90.73%	80.89%	97.24%	95.10%	88.50%
							Dentin	93.12%	96.82%	92.00%	78.46%	98.97%	91.84%	95.23%	90.83%	75.59%	98.46%

Figure 6 shows training and testing confusion matrices for the CNN model having three convolutional layers with 32, 256, 64 filters and three fully connected layers which provided the best results in the second study.

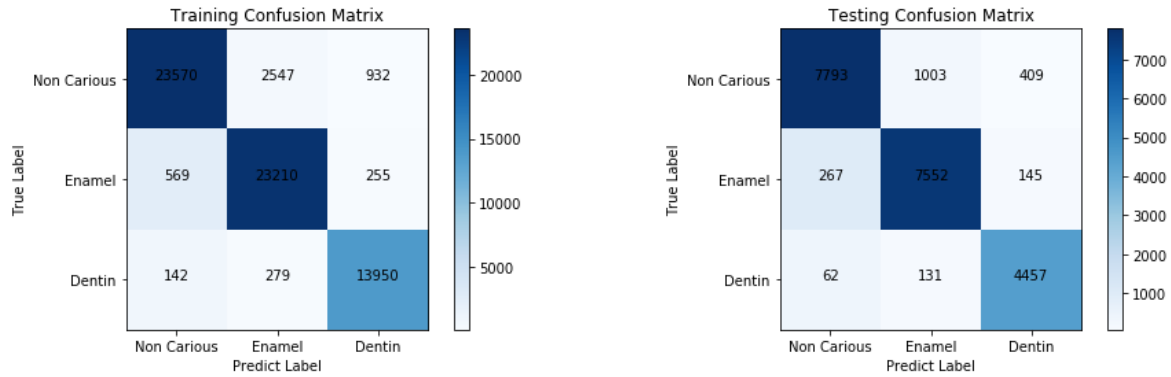


Figure 6 Training and testing confusion matrices for CNN model with three convolutional and three fully connected layers.

4. CONCLUSIONS

In this paper, human tooth specimens were imaged ex vivo using OCT imaging systems. A deep learning-based grading system for detection and classification of carious lesions was developed using CNN models. To investigate the CNN structures and determine the optimal numbers of convolutional and fully connected layers for OCT image classification tasks, we performed two studies. In the first study, the CNN models had two fully connected (FC) layers and three convolutional layers with five combinations obtained by changing the number of filters in the last convolutional layer of the feature extraction part of the CNN model. In the second study, the CNN models had three fully connected (FC) layers and three convolutional layers with five combinations. In each study, the diagnostic accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated for detection and diagnostic performances of the CNN models. This study is a step forward in the development of automated deep learning/OCT imaging system for early dental caries diagnosis, and could be extremely valuable in clinical studies and applications.

5. ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (NSF MRI-1920345) grant.

REFERENCES

- [1] L. Karlsson "Caries detection method based on changes in optical properties between healthy and carious tissue," *Int J Dent* 2010, 1-9 (2010).
- [2] JDB Featherstone, "Dental caries: a dynamic disease process," *Austr Dent J* 53 (3), 286-291 (2008).
- [3] R. H. Selwitz, A. I. Ismail, and N. B. Pitts, "Dental caries," *The Lancet* 369 (9555), 51-59 (2007).
- [4] EZ. Alsayed, I Hariri, A Sadr, et al. "Optical coherence tomography for evaluation of enamel and protective coatings," *Dent Mater J.* 34:98-107 (2015).
- [5] Y. Nakajima, Y Shimada, A Sadr, et al. "Detection of occlusal caries in primary teeth using swept source optical coherence tomography," *J. Biomed Opt.* 19:16020 (2014).
- [6] Y. Shimada, H Nakagawa, A Sadr, et al. "Noninvasive crosssectional imaging of proximal caries using swept-source optical coherence tomography (SS-OCT) in vivo," *J. Biophotonic* 7:506-513 (2014).
- [7] I. Wada, Y Shimada, M Ikeda, et al. "Clinical assessment of noncarious cervical lesion using swept-source optical coherence tomography," *J. Biophotonics* 8:846-854 (2014).
- [8] K. Imai, Y Shimada, A Sadr, Y Sumi, and J Tagami, "Noninvasive cross-sectional visualization of enamel cracks by optical coherence tomography in vitro," *J. Endod.* 38:1269-1274 (2012).

- [9] YC. Ahn, J Chung, P Wilder-Smith, and Z Chen, "Multimodality approach to optical early detection and mapping of oral neoplasia," *J. Biomed Opt.* 16:076007 (2011).
- [10] M. DeCoro, and P Wilder-Smith, "Potential of optical coherence tomography for early diagnosis of oral malignancies," *Expert Rev Anticancer Ther.* 10:321-329 (2010).
- [11] H. Kawakami-Wong, S Gu, MJ Hammer-Wilson, JB Epstein, Z Chen, and P Wilder-Smith, "In vivo optical coherence tomography based scoring of oral mucositis in human subjects: a pilot study," *J. Biomed Opt.* 12:051702 (2007).
- [12] ES. Matheny, NM Hanna, WG Jung, et al. "Optical coherence tomography of malignancy in hamster cheek pouches," *J. Biomed Opt.* 9:978-981 (2004).
- [13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* 521(7553), 436-444 (2015).
- [14] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science* 313(5786), 504-507 (2006).
- [15] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.* 18(7), 1527-1554 (2006).
- [16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Proceedings of the Advances in Neural Information Processing Systems, (MIT Press, 2012), pp. 1097-1105.
- [17] S.A. Kamran, S. Saha, A.S. Sabbir, A. Tavakkoli, Optic-net: a novel convolutional neural network for diagnosis of retinal diseases from optical tomography images, in 2019 IEEE International Conference On Machine Learning And Applications (ICMLA), 964-971 (2019).
- [18] Esteva, A., Chou, K., Yeung, S. et al. Deep learning-enabled medical computer vision. *npj Digit. Med.* 4, 5 (2021).
- [19] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentís, R. Zwigelaar, A. K. Davison, and R. Martí, "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks," *IEEE Journal of Biomedical and Health Informatics* 22(4), 1218-1226 (2018).
- [20] S. P. K. Karri, D. Chakraborty, and J. Chatterjee, "Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration," *Biomed. Opt. Express* 8(2), 579-592 (2017).
- [21] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *J. Bio. Optics Express* 8(5), 2732-2744 (2017).
- [22] N. Karimian, H. S. Salehi, M. Mahdian, H. Alnajjar, and A. Tadinada, "Deep learning classifier with optical coherence tomography images for early dental caries detection," *Proc. SPIE BIOS 10473, Lasers in Dentistry XXIV*, 1047304 (2018).
- [23] H. S. Salehi, M. Mahdian, M. M. Murshid, S. Judex, and A. Tadinada, "Deep learning-based quantitative analysis of dental caries using optical coherence tomography: an ex vivo study," *Proc. SPIE BIOS 10857, Lasers in Dentistry XXV*, 108570H (2019).
- [24] H. S. Salehi, M. Barchini, and M. Mahdian, "Optimization methods for deep neural networks classifying OCT images to detect dental caries," *Proc. SPIE BIOS 11217, Lasers in Dentistry XXVI*, 112170G (2020).