2023-05

CaViT: Early Stage Dental Caries Detection from Smartphone-image using Vision Transformer

Hossain, Md Shakhawat

Independent University, Bangladesh

Draft Version

CaViT: Early Stage Dental Caries Detection from Smartphone-image using Vision Transformer*

Md Shakhawat Hossain
RIoT Center, Dept. of CSE
Independent University, Bangladesh
Dhaka, Bangladesh
mshimul86sets@iub.edu.bd

Md Mahmudur Rahman
Dept. of CS
American International University Bangladesh
Dhaka, Bangladesh
mahmudur@aiub.edu

M M Mahbubul Syeed
RIoT Center, Dept. of CSE
Independent University, Bangladesh
Dhaka, Bangladesh
mahbubul.syeed@iub.edu.bd

Ummae Hamida Hannan
Dept. of CS
American International University Bangladesh
Dhaka, Bangladesh
21-91951-1@student.aiub.edu

Mohammad Faisal Uddin
RIoT Center, Dept. of CSE
Independent University, Bangladesh
Dhaka, Bangladesh
faisal@iub.edu.bd

Sahria Bakar Mumu
Universal Medical College Hospital
Dhaka, Bangladesh
sahriamumu41@gmail.com

***Abstract*—Caries detection is a routine clinical task in dental practice. If caries is detected at an early stage, non-invasive or micro-invasive treatment such as fillings and a root canal can be effective and thereby invasive treatment and therapies such as gum surgery and dental implants can be avoided. Invasive treatments are expensive and inappropriate for patients with low blood cell counts, cardiac problems and others health issues. Consequently, early caries detection is critical in dentistry. Caries is typically identified through a visual tactile examination in support of radiographic imaging. Fluorescence imaging, cone beam computed tomography or optical coherence tomography are also used. However, these procedures are time-consuming and expensive and require a physical examination of the patient. Moreover, the COVID-19 lessons taught us that such diagnosis should be avoided to prevent contagious diseases. Existing automated caries detection methods fail to achieve sufficient accuracy. Therefore, in this paper, we propose a highly accurate automatic system to detect early caries without any face-to-face interaction with the patient. This system is economical, rapid and easy to use. The proposed system uses a smartphone to capture teeth images and then relies on a vision transformer (ViT) to classify the images as advanced, early or no caries. Finally, the caries is segmented using a U-Net network. The proposed method outperformed the existing methods and achieved a sensitivity of 95%, 91% and 100% for the no caries, early caries and advanced caries classes when tested on a dataset of 300 images, developed for this study.**

***Index Terms*—dental caries, early caries detection, vision transformer, machine learning, smartphone image**

I. Introduction

Dental caries, also known as a cavity or tooth decay, is one of the most common health issues worldwide, second only to the common cold. It is an infectious process caused by acid-producing bacteria and is often associated with diets rich in sugars and refined carbohydrates. Primarily caries causes pain, if remains untreated it can lead to teeth loss, and heart disease and in extreme cases, it can cause death. If this bacterial infection happens in the upper back tooth it can spread to the sinus behind the eye and then to the brain which causes death. However, if caries is detected at an early stage then it can be cured and serious consequences can be avoided. For instance, caries that only affect the enamel or the outer layers of the dentin could be cured or even reversed. Caries starts with demineralizing the surface of the teeth due to bacterial plaque. In the early stage, the demineralization remains limited to enamel, dentin or cementum commonly called hard tissues of the tooth. If it is detected at this stage, non-invasive treatment can be provided to remineralize the surface to maximize the retention of tooth tissue and save the patient from the lifelong cycle of tooth restoration. On the other hand, if it advances into the deep dentine or pulp, breaking the tooth's surface, then it will need restoration treatment because it is no longer possible to reverse the impact. This signifies the importance of early caries detection. Fig. 1 shows the anatomy of a tooth.

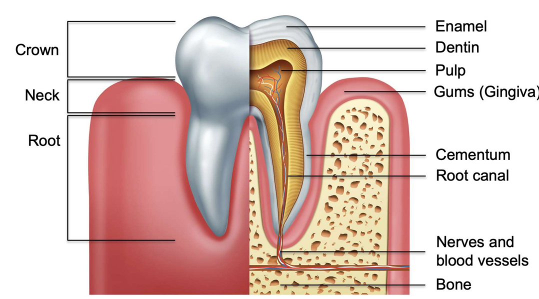

Fig. 1. Anatomy of a tooth.

In the United States, more than 90% of adults experience dental caries before they reach 30 years of age [1]. This number is around 82.7% for Bangladesh according to a study [2]. The same study also reports that the prevalence is significantly higher in rural, low-income, and illiterate families, particularly in children between the ages of 8 and 10. Additionally, due to the cost of dental examinations and the lack of financial

This research is funded by RIoT Research Center.

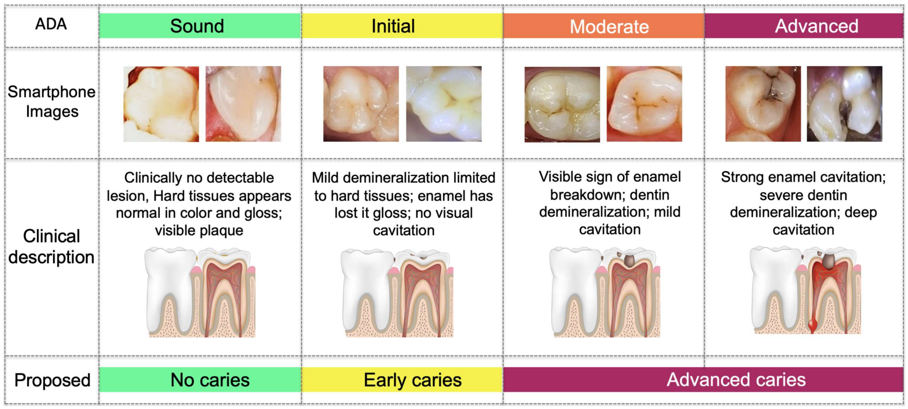

Fig. 2. Taxonomy of dental caries.

aid or insurance, most people don't consult a dentist or visit a clinic unless the condition is severe. Moreover, the number of dentist is also limited and a patient has to follow a long queue for a consultation that lasts less than five minutes [3]. Therefore, it is necessary to develop a cheap and easy-to-use automatic early caries detection method that can don't require a face-to-face interaction.

Several automatic caries detection methods have been proposed previously. These systems mainly utilize periapical, panoramic x-ray, bitewing radiography or visible-band RGB image. Haihua et al. [4] and Lian et al. [5] proposed automatic caries classification methods using panoramic x-ray images. Shinae et al. [6] utilized bitewing intra-oral radiography image and classified caries into initial, moderate and extensive classes. These methods require the patient to visit a dentist to obtain a radiography image, which is then processed by an algorithm to detect caries. Although some of these technologies have attained acceptable accuracy, they are time-consuming and expensive. In this work, we also intend to develop an affordable and fast detection system that does not rely on the availability of an expert. Methods based on visible-band imaging are thus appropriate for this purpose. Methods proposed by Datta et al. [7] and Koutsouri et al. [8] utilized RGB images captured by digital camera to detect caries. These methods perform binary classification on the RGB images of tooth. Datta et al. detected caries with 83.63% sensitivity, while Koutsouri et al. detected it with 92% sensitivity. These system achieved binary classification for caries and didn't differentiate between the stages of caries. Another binary caries detection method was proposed by Jiang et al. [9]. This method used a smartphone-captured image, although it only obtained a sensitivity of 60.4%. Smartphones are easily available and simple to use. As a result, smartphone-based technologies are more suitable for the proposed system. Automated caries detection methods using smart-phone captured images were proposed by Thanh et al. [10] and Duong et al. [11]. These methods incorporated early-stage caries detection and classified caries into three stages. However, the accuracy of early-stage caries detection was low. These methods mainly fail to different early caries from healthy teeth. Thanh et al. [10] detected early caries with only 36.9% sensitivity while Duong et al. [11] achieved 88.1% sensitivity. This indicates the need for a more precise smartphone-based early caries detection system. In this paper, we have proposed an automatic system that can detect caries at early stage from smartphone captured tooth image. This system relies on the user to upload an image of a tooth and then classify it utilizing vision transformer based deep learning model with an accuray of 95% approximately. American Dental Association (ADA) classy cares as sound, initial, moderate or advanced [12]. The proposed system adopted the taxonomy and simplified to classify caries as no caries, early caries or advance caries, as illustrated in Fig. 2.

The major contributions of this paper are: 1) development of an automated early caries detection method that don't require face-to-face interaction, 2) development of a highly accurate, cheap, rapid and easy-to-use caries detection method, 3) development of a smartphone captured caries image dataset and 4) Dentists' evaluation of the proposed method for practical use.

II. Materials and Methodology

Algorithm 1 shows the steps of the proposed system. The proposed system starts with capturing an image of a tooth using a smartphone. Then the image is evaluated to ensure its quality and color distribution. Smartphone-captured images are highly vulnerable to focus blur errors, therefore, we estimated the quality and rejected poor-quality images. For the quality evaluation, we utilized the reference-less method proposed by Yagi et. al. [13] and optimized it for the smartphone-captured image by updating the α, β and γ coefficients of predictors as given in the linear regression model given in Eq. 1:

$$Quality = \alpha + \beta \times blurriness + \gamma \times noise \quad (1)$$

After that, the color values of the good-quality images are normalized as the color varies significantly depending on the smartphone used to capture the image. Then the color-normalized image is resized and classified using the fine-tuned

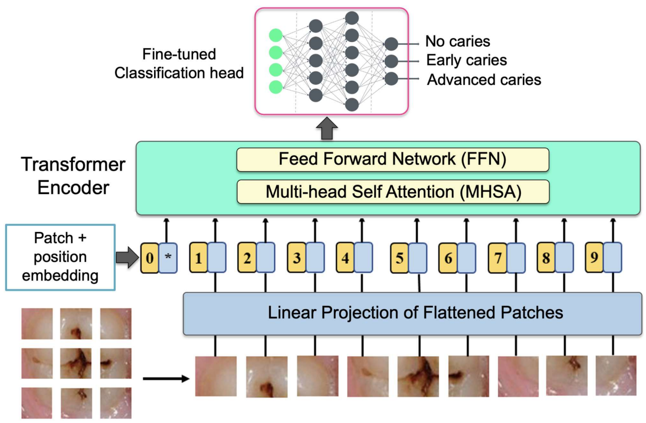

Fig. 3. Architecture of the Vision Transformer for caries classification.

ViT model. If the ViT classified the image as Early caries or Advanced caries only then the image is processed using the U-Net segmentation model to segment caries. Finally, the system returns the caries classification and localization information. The design of this architecture was motivated by the existing computer vision systems [14], [15] to locate and classify small objects from images. The details of the dataset, caries classification and segmentation model are given in the following sections.

A. Dataset

In this study, we prepared a dataset of tooth images to train, validate and test the machine-learning models. A total of 1200 images were collected from multiple dental clinics in Bangladesh. This dataset included 400 no caries, 400 early caries and 400 advanced caries images. The images were captured by different smartphones which include iPhone 6, iPhone 10, Realme, Samsung and Vivo. Ethical approval was obtained from the Institutional Review Board at the Independent University, Bangladesh for this research (approval code: 2022-SETS-002).

B. Caries Classification using ViT

Transformer-based model is possibly the first machine-learning architecture that became dominant in two different areas: natural language processing and computer vision. The proposed method utilized transformer-based models to classify caries. ViT has recently gained immense popularity as an alternative to convolution neural network (CNN) [16] in computer vision. In comparison to CNN, ViT shows a generally weaker inductive bias resulting in increased reliance on model regularization or data augmentation when training on smaller datasets. To evaluate the performance of the ViT model for caries classification, we took a pre-trained ViT model on ImageNet dataset [17] with 1000 classes and optimized it in three different approaches: 1) used it as a feature extractor (FE) 2) fine-tuned only multi-head self-attention layers (MHSA) of transformer 3) fine-tuned the MHSA layers and the customized the classification (MCTN) head using our dataset.

The ViT model divides the image into fixed-size patches. Then these patches are flattened and combined with position embeddings to a sequence that is fed to the transformer encoder. The transformer encoder consists of multiple blocks each of which contains normalization, MHSA and multi-layer perceptron layers (MLP). The output of the transformer encoder is fed to a classification head that consists of MLP to map the encoded feature vector to output classes.

Algorithm 1 Early caries detection

Input: I_{STI}, Q_{th}
  I_{STI}: smartphone captured tooth images
  Q_{th}: quality threshold
  ViT: vision transformer with parameters
  $SegNet$: caries segmentation network with parameters
Initialisation:
  $Q_{th} \leftarrow 5$
while $I_{STI} != NIL$ **do**
  Estimate $I_{Quality}$ for I_{STI} using Eq. 1
  if $I_{Quality} \leq Q_{th}$ **then**
    Estimate C_{Linear} from I_{STI}
    if $C_{Linear} \leq 0.0031$ **then**
      $I_{sRGB} = 12.92 \times C_{Linear}$
    else
      $I_{sRGB} = 1.0552 \times C_{Linear}^{\frac{1}{2.4}}$
    end if
    $I_R \leftarrow$ Resize I_{sRGB} to 224 x 224
    $ViT_{out} = ViT(I_{sRGB})$
    if ViT_{out}==1 **then**
      $\psi \leftarrow$ Advanced caries
    else if ViT_{out}==2 **then**
      $\psi \leftarrow$ Early caries
    else
      $\psi \leftarrow$ No caries
    end if
    if ViT_{out}==1 $\|$ViT$_{out}$==2 **then**
      $Seg_{map} = SegNet(I_{sRGB})$
    end if
  end if
end while
return ψ, Seg_{map}

In the first approach, a pre-trained ViT model was simply used for caries classification. In the second approach, the MHSA layers were fine-tuned using our dataset. MHSA allows the network to control the mixing of information between parts of an input sequence which leads to a better representation to increase the performance of the model. Finally, in the third approach, we added more dense layers to gradually reduce the output layers to three neurons. Then fine-tuned the customized classification head along with the MHSA layers. The architecture of the third approach is shown in Fig. 3 which achieved the highest performance in our experiment.

The ImageNet dataset has 1000 classes thus the final layer of the ViT has 1000 output neuron while the proposed system requires only 3 output neurons in the final layer. In our

TABLE I
TOP 15 NETWORKS FINE-TUNED FOR CARIES CLASSIFICATION.

Rank	Arch. Name	Final layers	Optimizer	Train acc	Train loss	Val acc	Val loss	Test acc
1	$MCTN_{13}$	$1000-128-64-32-3$	AdamW	0.983	0.045	0.940	0.126	0.953
2	$MCTN_{11}$	$1000-128-32-3$	AdamW	0.976	0.060	0.911	0.327	0.945
3	$MCTN_{10}$	$1000-64-32-3$	AdamW	0.965	0.085	0.937	0.192	0.937
4	$MCTN_{6}$	$1000-8-3$	AdamW	0.887	0.297	0.854	0.476	0.937
5	$MCTN_{9}$	$1000-128-32-8-3$	AdamW	0.943	0.179	0.859	0.476	0.935
6	$MCTN_{2}$	$1000-128-64-32-16-8-3$	AdamW	0.981	0.064	0.911	0.037	0.929
7	$MCTN_{12}$	$1000-128-64-3$	AdamW	0.939	0.195	0.929	0.226	0.929
8	$MCTN_{3}$	$1000-64-32-16-8-3$	AdamW	0.985	0.042	0.901	0.493	0.921
9	$MCTN_{8}$	$1000-256-64-16-3$	AdamW	0.981	0.056	0.932	0.350	0.906
10	$MCTN_{4}$	$1000-32-16-8-3$	AdamW	0.915	0.243	0.885	0.371	0.867
11	$MCTN_{5}$	$1000-16-8-3$	AdamW	0.927	0.209	0.859	0.389	0.859
12	$MCTN_{7}$	$1000-3$	AdamW	0.842	0.395	0.836	0.428	0.828
13	$MCTN_{1}$	$1000-256-128-64-32-16-8-3$	AdamW	0.982	0.042	0.921	0.045	0.820
14	$MCTN_{19}$	$1000-256-32-16-8-3$	AdamW	0.973	0.142	0.920	0.245	0.813
15	$MCTN_{16}$	$1000-256-64-16-8-3$	AdamW	0.922	0.194	0.891	0.270	0.813

experiment, we have found that if the neurons are reduced to 3 directly from 1000 then it effects the accuracy. Therefore, we gradually reduced the neuron using different steps which created a set of 13 architectures for the third approach. Each architecture were then trained and validated using 80% and 20% of the 900 images, accordingly which included 300 images for each class. Another 300 images consists of 100 images from each class were used for testing the models.

In this study, we used "$ViT-B/32$" models for all three approaches which is the Base variant of vision transformer with 32×32 input patch size. These models were trained for 250 epochs using two different optimizer AdamW and SGD and sparse categorical cross entropy loss function. Then, all the networks from the three approaches were compared and the network with highest test accuracy was selected for the proposed system which is the $MCTN_{13}$ network derived from the third approach.

C. Caries Segmentation using ResU-Net++

After the caries classification, caries are segmented from the image if it is classified as early caries or advanced caries. To segment caries, we utilized ResUnet++ [18] segmentation network. ResUNet++ is an advancement of vanilla U-Net and original ResUNet models to achieve better segmentation performance for medical image analysis. Therefore, we used this model and optimized it for our work using transfer learning approach. To train and test the network 300 images were annotated by two experts. As the network is designed for semantic segmentation, the images were annotated to create binary masks in which the white pixel indicates caries. This image-set included 150 early carries and 150 advanced caries images. The model was trained using 200 images and tested using the rest.

During the training, data augmentation was applied using vertical flip, horizontal flip and rotating the images at 15 degrees intervals starting from 15 to 360 degrees. The network was trained for 50 epochs using a 0.0001 learning rate, Adam optimizer and binary cross-entropy loss function.

III. RESULTS

A. Classification Result

We compared networks trained with the three techniques, FE, MHSA, and MCTN. Then they were ranked based on their accuracy on the test dataset. The networks were trained, validated, and tested using the same dataset. The top 15 networks based on test accuracy are shown in Table I. It demonstrates that MCTN-based networks outperform FE and MHSA-based networks. Based on the results of this experiment, it can be concluded that fine-tuning the final classification layers in addition to the attention layers enhances classification accuracy more than fine-tuning the attention layers alone. Due to time and memory constraints, we trained only the MSHA and final classification layers in this experiment rather than the whole network. The AdamW optimizer produced the best results for all models. Fig. 4 shows the loss and accuracy curves for training and validating the top network $MCTN13$ which achieved a test accuracy of 95.3%. Therefore, $MCTN13$ is selected for the proposed system for caries classification. Fig. 5 shows the confusion matrix for the proposed caries classification method on the test dataset. The proposed caries classification method achieved an accuracy of 95.3%. The precision values were 91.3%, 94.7% and 100% for the no caries, early caries and advanced caries classes, accordingly. The sensitivity for no caries, early caries and advanced caries were 95%, 91% and 100% accordingly. Finally, we compared the results of the proposed method with the existing methods in the literature, shown in Table II. The proposed method outperformed the existing method in terms of sensitivity, precision and accuracy.

B. Segmentation Result

After the caries classification, the early and advanced caries were segmented in the proposed system using the trained ResUNet++ model, as illustrated in Fig. 6. The segmentation model achieved a mean intersection over union (IoU) of 0.98 for the test dataset. The IoU measures the number of pixels common between the annotation mask prepared manually by the experts and the ResU-net++ prediction mask divided by

TABLE II
COMPARISON OF PROPOSED METHOD WITH THE EXISTING METHODS.

Methods	Imaging modality	Classes	Sensitivity
Thanh et al.	Smartphone image	Non-cavitated, cavitated,late-cavitated	Non-cavitated vs cavitated: 36.9%, Cavitated vs late-cavitated: 87.4%
Shinae et al.	Bitewing radiographs	Initial, Moderate, Extensive	Initial: 74.7%, Moderate: 90.5%, Extensive: 95.4%
Haihua Zhu1 et al.	Panoramic X-ray	Shallow caries, Moderate caries, Deep caries	Shallow caries: 93%, Moderate caries: 69.4%, Deep caries: 97.1%
Luya Lian et al.	Panoramic X-ray	D1: Enamel or outer dentin caries, D2: Middle dentin caries, D3: Inner dentin or pulp caries	D1: 76.5%, D2: 65.2%, D3: 91.8%
Hao Jiang et al.	Smartphone image	Tooth decay (Yes/No)	60.4%
Duong et al.	Smartphone image	No surface change, Non-Cavitated, Cavitated	No surface change vs non-cavitated: 88.1%, Non-Cavitated vs cavitated: 88.2%
Soma Datta et al.	Digital RGB image	Caries (Yes/No)	83.63%
Koutsouri et al.	Digital RGB image	Caries (Yes/No)	92%
Proposed	Smartphone image	No caries, Early caries, Advance caries	No caries: 95.3%, Early caries: 91%, Advance caries: 100%

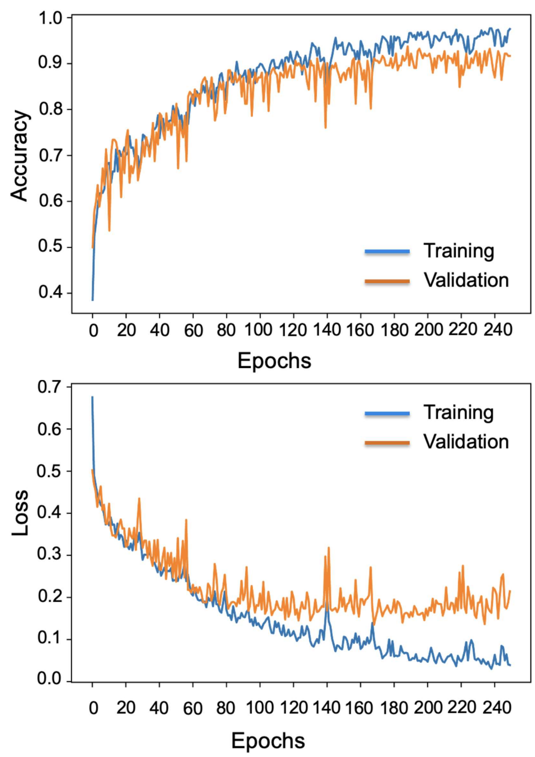

Fig. 4. Training and validation curves for selected model for caries classification.

the total number of pixels covered by both for an image, as given in Eq. 2.

$$IoU = \frac{Annotation \cap Prediction}{Annotation \cup Prediction} \tag{2}$$

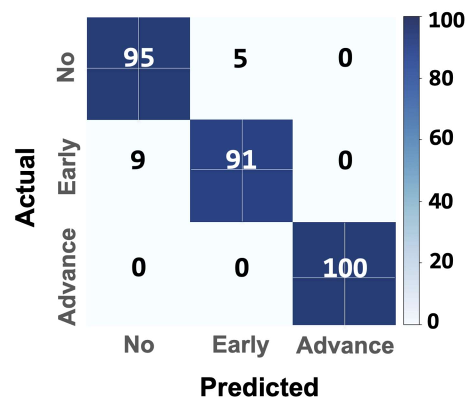

Fig. 5. Confusion matrix of test data for caries classification .

TABLE III
EVALUATION OF RESUNET++ MODEL FOR CARIES SEGMENTATION.

Metrics	$ResUNet{+}{+}$
Training accuracy	0.995
Training loss	0.016
Training mean IoU	0.989
Validation accuracy	0.993
Validation loss	0.061
Validation mean IoU	0.987
Test mean IoU	0.980

C. Evaluation for the practical use

The practical usability of the proposed system was evaluated in terms of accuracy, cost, time and user-friendliness. The proposed method outperformed the existing automatic caries detection methods. We developed an Android application to implement the proposed method. The application was then evaluated by two experts and two individual testers to ensure its practical use. In the evaluation, this application took less than one minute to get the caries classification result after

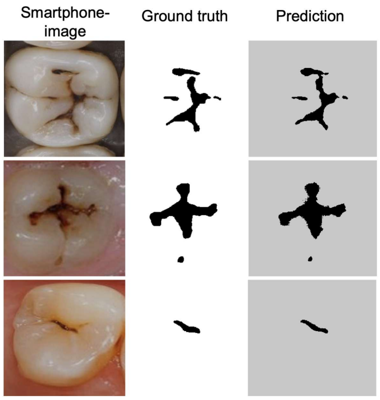

Fig. 6. Caries segmentation by proposed method (color of the ground truth and prediction image are modified for visibility).

uploading the image to the application. This demonstrates the system's rapid caries classification ability. This system relies on a smartphone to capture an image and launch the application. Thus, it is less expensive than alternative systems that use radiography or similar imaging technology. To make the system easier to use, we incorporated comments of the experts and individual testers. The proposed system doesn't require a patient to meet a technician to capture the tooth image or a dentist for consultation. This type of system useful to combat COVID-19 and other infectious diseases.

IV. Conclusion

In this paper, we propose an accurate, easy-to-use, cheap and rapid caries detection system. Detecting caries at an early stage is critical as it doesn't show distinguishable visible features compared to a healthy tooth. From Table II and Fig. 5 it can be observed that differentiating the early stage caries is the most difficult task for automatic caries detection. Although, the proposed method achieved an accuracy of 91% for detecting early caries which is significantly higher than the previous methods. Most of the methods achieved a high accuracy in detecting advanced caries. The proposed method detected advanced caries at 100% accuracy and sensitivity. The proposed method relies on the smartphone-captured image which is not harmful like x-ray imaging. This type of system can assist dentists to reduce the examination time of a patient. One of the limitations of this system it expects the patient to upload an image containing a single tooth. Therefore, it is necessary to evaluate the performance of the system if multiple teeth exist in the image with a different type of caries.

Acknowledgment

We express our gratitude and appreciation to Dr. Dipto Ghosh, Dr. Sahria Bakar Mumu and Uamme Hamida Hannan for thier contribution to develop and annotate the dataset. We are also thankful to RIoT for funding this project.

References

[1] Dye, B., Tan, S., Smith, V., Barker, L., Thornton-Evans, G., Eke, P. & Beltrán-Aguilar, E. Trends in oral health status; United States, 1988-1994 and 1999-2004. (2007)

[2] Sultana, S., Parvin, M., Islam, M., Chowdhury, E. & Bari, A. Prevalence of Dental Caries in Children in Mymensingh and Its Associated Risk Factors: A Cross-Sectional Study. *Dentistry Journal*. **10**, 138 (2022)

[3] Hossain, M., Syeed, M., Fatema, K. & Uddin, M. The Perception of Health Professionals in Bangladesh toward the Digitalization of the Health Sector. *International Journal Of Environmental Research And Public Health*. **19**, 13695 (2022)

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

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

[6] Lee, S., Oh, S., Jo, J., Kang, S., Shin, Y. & Park, J. Deep learning for early dental caries detection in bitewing radiographs. *Scientific Reports*. **11**, 16807 (2021)

[7] Datta, S. & Chaki, N. Detection of dental caries lesion at early stage based on image analysis technique. *2015 IEEE International Conference On Computer Graphics, Vision And Information Security (CGVIS)*. pp. 89-93 (2015)

[8] Koutsouri, G., Berdouses, E., Tripoliti, E., Oulis, C. & Fotiadis, D. Detection of occlusal caries based on digital image processing. *13th IEEE International Conference On BioInformatics And BioEngineering*. pp. 1-4 (2013)

[9] Jiang, H., Zhang, P., Che, C., Jin, B. & Others Rdfnet: A fast caries detection method incorporating transformer mechanism. *Computational And Mathematical Methods In Medicine*. **2021** (2021)

[10] Thanh, M., Van Toan, N., Ngoc, V., Tra, N., Giap, C. & Nguyen, D. Deep learning application in dental caries detection using intraoral photos taken by smartphones. *Applied Sciences*. **12**, 5504 (2022)

[11] Duong, D., Kabir, M. & Kuo, R. Automated caries detection with smartphone color photography using machine learning. *Health Informatics Journal*. **27**, 14604582211007530 (2021)

[12] Young, D., Nový, B., Zeller, G., Hale, R., Hart, T., Truelove, E., Ekstrand, K., Featherstone, J., Fontana, M., Ismail, A. & Others The American Dental Association caries classification system for clinical practice: a report of the American Dental Association Council on Scientific Affairs. *The Journal Of The American Dental Association*. **146**, 79-86 (2015)

[13] Hossain, M., Nakamura, T., Kimura, F., Yagi, Y. & Yamaguchi, M. Practical image quality evaluation for whole slide imaging scanner. *Biomedical Imaging And Sensing Conference*. **10711** pp. 203-206 (2018)

[14] Hossain, M., Raihan, M., Hossain, M., Syeed, M., Rashid, H. & Reza, M. Aedes Larva Detection Using Ensemble Learning to Prevent Dengue Endemic. *BioMedInformatics*. **2**, 405-423 (2022)

[15] Hossain, M., Shahriar, G., Syeed, M., Uddin, M., Hasan, M., Hossain, M. & Bari, R. Tissue Artifact Segmentation and Severity Assessment for Automatic Analysis Using WSI. *IEEE Access*. **11** pp. 21977-21991 (2023)

[16] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. & Others An image is worth 16x16 words: Transformers for image recognition at scale. *ArXiv Preprint ArXiv:2010.11929*. (2020)

[17] Deng, J., Dong, W., Socher, R., Li, L., Li, K. & Fei-Fei, L. Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference On Computer Vision And Pattern Recognition*. pp. 248-255 (2009)

[18] Jha, D., Smedsrud, P., Riegler, M., Johansen, D., De Lange, T., Halvorsen, P. & Johansen, H. Resunet++: An advanced architecture for medical image segmentation. *2019 IEEE International Symposium On Multimedia (ISM)*. pp. 225-2255 (2019)