Automated Segmentation of Gingival Diseases from Oral Images

Aman Rana[1], Gregory Yauney[1], Lawrence C. Wong[2], Otkrist Gupta[1], Ali Muftu[2], Pratik Shah[1*]

[1] *MIT Media Lab*
Massachusetts Institute of Technology
Cambridge, MA, USA
{*arana, gyauney, otkrist, pratiks*}*@mit.edu*

[2] *School of Dental Medicine*
Tufts University
Boston, MA
{*lawrence.wong, ali.muftu*}*@tufts.edu*

***Abstract*— Periodontal diseases are the largest cause of tooth loss among people of all ages and are also correlated with systemic diseases such as endocarditis. Advanced periodontal disease comprises degradation of surrounding tooth structures, severe inflammation and gingival bleeding. Inflammation is an early indicator of periodontal disease. Early detection and preventive measures can help prevent serious occurrences of periodontal diseases and in most cases restore oral health. We report a machine learning classifier, trained with annotations from dental professionals, that successfully provides pixel-wise inflammation segmentations of color-augmented intraoral images. The classifier successfully distinguishes between inflamed and healthy gingiva and its area under the receiver operating characteristic curve is 0.746, with precision and recall of 0.347 and 0.621 respectively. Dental professionals and patients can benefit from automated point-of-care early diagnosis of periodontal diseases provided by this classifier using oral images acquired by intraoral imaging devices.**

Keywords: biomarkers, periodontal diseases, gingivitis, convolutional neural networks, deep learning, image segmentation, dentistry

I. INTRODUCTION

Gingivitis is the inflammation of gingiva around the tooth, making the gums sensitive and likely to bleed. Gingivitis can progress and lead to periodontitis, with severe inflammation and infections in the surrounding structures of the teeth. If left untreated, periodontal diseases can cause progressive bone destruction and ultimately loss of tooth. Early detection and treatment helps treat gingivitis and prevent tooth loss [1][2].

Biomarkers provide a fast, accurate and non-invasive way to diagnose several diseases and thus can be used for prognostic screening along with monitoring of clinical responses. Porphyin biomarkers have been used to detect dental plaque and inflammation of gingival surfaces. Gums and teeth, when illuminated with a blue light (405-450 nm), fluoresce in the presence of porphyrin, produced by oral bacteria in plaque biofilms and present in hemoglobin in the blood [3]. Gingivitis results in an increased blood flow around the inflamed gingiva, leading in turn to increased red fluorescence (650 nm wavelength) from porphyrin in the surrounding vasculature. Diagnoses of gingival diseases using intraoral imaging technologies is precluded due to the high cost of specialized imaging systems and lack of automated diagnoses.

Visual inspection and probing techniques have been traditionally used for diagnosis of gingival inflammation in patients [4]. Although accurate, these methods are considered subjective due to differences in training, experience and location of the hygienists and dentists, creating errors in early diagnosis of gingivitis. Computer vision, machine learning and deep neural networks are becoming more effective and can now perform automated and accurate diagnoses of several diseases [5]

We describe an automated system that performs pixel-wise segmentation of the inflamed gingiva to detect gingivitis and periodontal disease using fluorescence images acquired by an FDA-approved intraoral camera [6]. Intraoral fluorescent images from 150 consenting adults, aged 18-90 years old, were analyzed by dentists for gingivitis and then used to train a machine learning classifier. The trained classifier accepts an intraoral image of gums and teeth and provides a localized and automated detection of gingival inflammation and periodontal disease on a per-pixel basis.

II. RELATED WORK

A. Periodontal disease detection

Juan et. al used computer vision techniques and incorporated an off-the-shelf camera to automatically predict gingival probe depth using training data with ground truth measurements [7]. The camera is attached to the probe and depth predictions are made with reasonable accuracy. To the best of our knowledge, this remains the only attempt to create an automated solution to estimate gingival disease using images and computer vision. This approach however lacked other key parameters such as inflammation and bleeding indices, hypervascularization and papillary margin quality and depends on a clinically invasive procedure to inform diagnoses.

B. Deep neural networks

Convolutional neural networks (CNNs) are deep neural network (DNN) architectures that learn to identify features from images [8]. The classifier learns using backpropagation

*Corresponding author.

978-1-5386-1392-4/17/$31.00 ©2017 IEEE

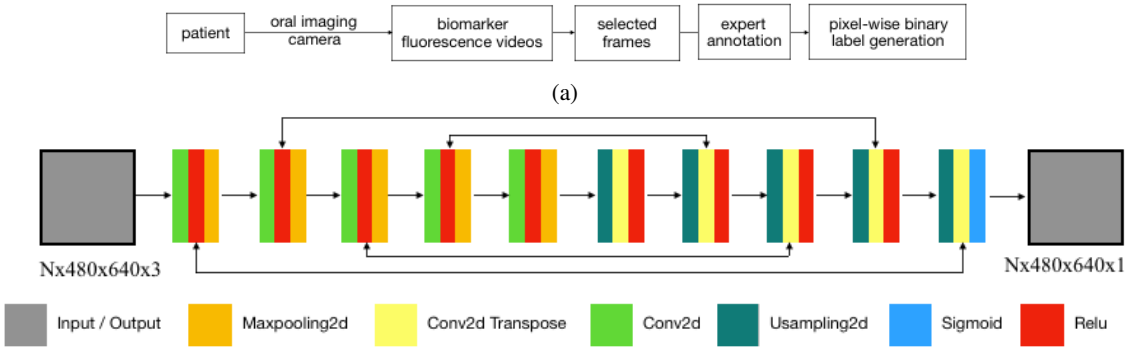


Fig. 1: (a) The data acquisition and labeling process. (b) Autoencoder network consisting of convolutional, maximum pooling, upsampling, rectified linear units and sigmoid layers. Arrows between non-adjacent layers represent skip connections. The network takes as input three-channel images of size 480×640×3 pixels and outputs a single-channel segmented image of size 480×640 pixels. N is the batch size used for one iteration of training.

until the network is trained for the specified task. Maximum pooling is used for translational invariance. CNNs are highly effective in identifying features in images with high accuracy, given enough training data [9] [10].

AutoEncoders (AE) combined with convolutional layers have been used for segmentation of input images [11].

III. Technical Approach

A. Data collection

Data was collected during a study at the 2015 Kumbh Mela in Nashik, India from consenting adults (aged 18-90 years) using ACTEON Soprocare, an FDA-approved oral imaging camera (ACTEON North America, Mount Laurel, New Jersey, USA). The Massachusetts Institute of Technology Committee on Humans as Experimental Subjects reviewed and approved protocol 1603518893 associated with this data. The venue provided an opportunity to collect medical data from a diverse group of people. The teeth were illuminated with light of 405-450 nm wavelength and the corresponding fluorescence captured using the oral imaging device. The final dataset used for this study comprises 405 color-augmented intraoral biomarker images from 150 individuals. The data acquisition process can be seen in Fig. 1.

The oral imaging device captures fluorescence from the biomarker porphyrin, an indicator of periodontal disease, and outputs color-augmented white light images. Plaque is displayed in shades of yellow and orange while gingival inflammation is displayed in shades of magenta and red [3].

B. Ground truth generation

A graphical interface was developed to display images to a dentist for annotations (Fig. 2). A dental expert provided bounding boxes around regions of inflamed gingiva along with a modified gingival index (MGI) between 0 and 4 (inclusive), for each image [12]. The boxes provided by the dental professional were representative, not exhaustive, resulting in some of the inflamed gingiva pixels lying outside the bounding boxes, which could lead to faulty training. A final bounding box around the annotated bounding boxes was calculated, which served to delineate the general area of gingival inflammation. The pixels inside the bounding box corresponding to inflamed gingiva (identified using intensity thresholding) were given a value of 1; rest of the pixels were given a value of 0. This operation resulted in 405 pairs of images and corresponding inflammation segmentations.

C. Data Preparation

The image dataset was divided into training and validation data: 258 images (63.7%) and 147 images (36.3%) respectively. The distribution of the dataset across MGI scores can be seen in Table I. Sampling with replacement was used to create training batches, where all values of MGI were equally represented, thereby preventing bias towards any specific value. Random vertical flipping was applied to the training images to induce rotational invariance to gingival positions. The validation dataset was randomly selected to preserve the proportion of images with each MGI and to preserve the variety of tooth types and photographed gingival positions.

D. Classifier model

The deep learning network architecture, shown in Fig. 1b, uses an autoencoder framework combined with convolu-

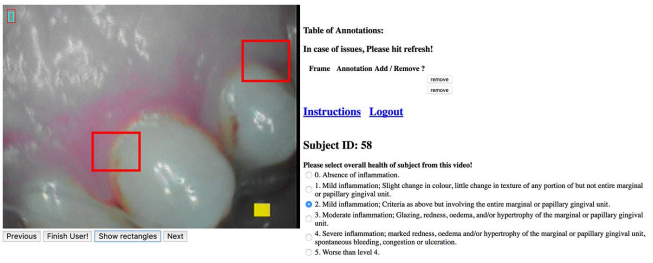


Fig. 2: Graphical interface used for annotation by dental professionals.

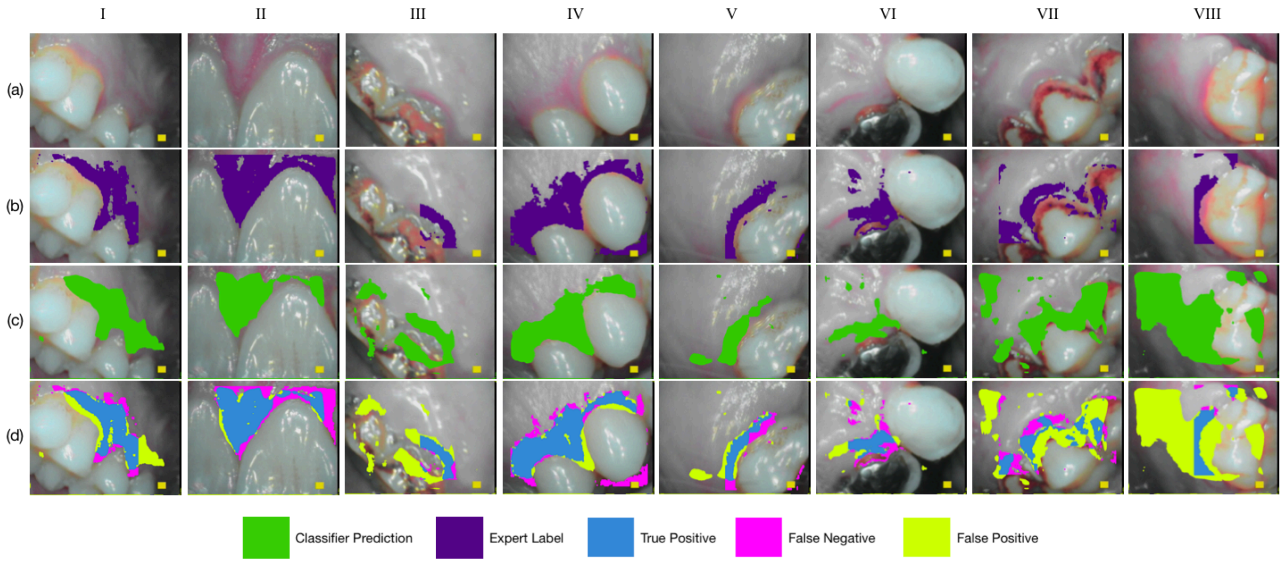


Fig. 3: Representative sets of images from the validation dataset. The columns (I-VIII) represent images from different patients. (a) Intra-oral images used as input to the classifier. (b) Associated ground truth localized labels constructed from expert bounding boxes and thresholding. (c) Output segmentations from the classifier. (d) Segmentation errors colored by type of error.

tional layers. The network consists of convolutional layers, maximum pooling layers, upsampling, rectified linear unit and final sigmoid activation. The architecture aims to learn the mapping between the input image and the ground-truth gingival inflammation segmentation. Residual connections have been utilized to prevent gradient vanishing and expedite learning [13]. The classifier takes the color-augmented RGB images with dimensions of 640×480 pixels as input and single channel binary image of size 640×480 as the ground truth. Dice loss was used as the loss function and the classifier is trained using adaptive gradient descent with momentum [14]. The classifier was implemented in TensorFlow and trained using a single NVIDIA GeForce GTX Titan X GPU [15]. A batch size of 32 was used along with an initial learning rate of 1×10^{-6}. The learning rate decreased by a factor of five every 500 iterations. Training was performed for 5000 iterations. The trained classifier accepts as input a color-augmented intraoral image and outputs a pixel-wise segmentation.

Modified gingival index	No. training images	No. validation images
1	47 (18.2%)	40 (27.2%)
2	133 (51.5%)	65 (44.2%)
3	61 (23.6%)	28 (19.1%)
4	17 (6.6%)	14 (9.5%)

TABLE I: Distribution of training and validation data. The modified gingival index values range from 0 (healthy gums) to 4 (advanced gingivitis with inflammation). No images were scored 0. Percentages were calculated with respect to the total number of images in the column's dataset.

IV. RESULTS AND DISCUSSION

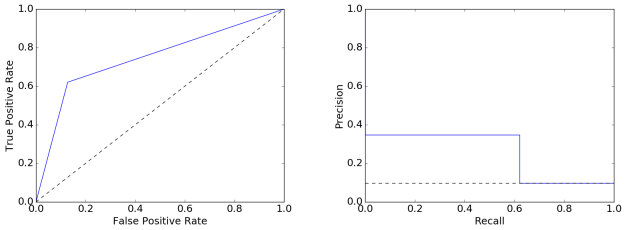
The classifier produces a pixel-wise segmentation of areas predicted to contain gingival inflammation. Row (c) in Fig. 3 shows the segmentations results for representative images in the validation set. Row (d) in Fig. 3 shows the errors for the segmentations in the validation dataset. Fig. 4a shows the receiver operating characteristic (ROC) curve for the validation set for pixel-level segmentations. The area under the ROC curve (AUC) was 0.746, indicating a 0.746 probability that pixels corresponding to inflamed gingiva were more likely to be included in the segmentations than pixels corresponding to non-inflamed gingiva. The precision and recall values are 0.347 and 0.621, respectively (Fig. 4b); 34.7% of all pixels classified as inflammation were actually inflamed, and 62.1% of all truly inflamed pixels are correctly classified as inflammation.

The classifier segmentation was validated by three dentists, and the agreement among the experts and between the classifier and each expert can be seen in Table II. The AUC between the classifier and each dentist averages to 0.7372, indicating that there is a 73.72% chance that the classifier classifies a pixel that a given dentist considers inflamed as more likely to be inflamed than a pixel the dentist does not consider inflamed. Likewise, the AUC between the dentists pairs averages to 0.7754 (77.54%). The agreement between the classifier and the dentists is similar, despite being trained on annotations from a single dentist (dentist C), indicating that the classifier is able to identify a general representation of gingivitis.

The trained classifier identified with high accuracy areas with gingivitis and periodontal diseases in validation images.

Source 1	Source 2	AUC
Dentist A	Dentist B	0.7525
Dentist B	Dentist C	0.7893
Dentist C	Dentist A	0.7844
Dentist A	Classifier	0.7357
Dentist B	Classifier	0.7300
Dentist C	Classifier	0.7460

TABLE II: Agreement among the dentists and between each dentist and classifier. Higher AUC indicates better agreement. AUC: area under the receiver operating characteristic curve.



(a) (b)

Fig. 4: (a) Receiver operating characteristics curve for the validation dataset. The area under the curve is 0.746. The black dashed line represents random chance. (b) Precision-recall curve for the validation dataset. The precision value for the classifier is 0.347 and the recall value is 0.621.

For example, comparing images in Fig. 3, IId showed that the segmentation from the classifier IIc successfully matched the ground truth labels provided by the expert IIb. Similar results were observed across the majority of other image pairs we validated. The low rates of false positives predicted by the classifier in majority of validation images also indicated that it does not segment out gingival inflammation in healthy tissue. While a low false negative rate indicated comprehensive recognition of surfaces associated with bona fide gingival inflammation. An exception was seen in panel VIIId where we observed errors primarily attributed to false positive segmentation in potentially healthy tissues.

The observed errors may be attributable to aggressive color augmentation of non-inflamed tissues by the intraoral camera for easy visualization vs. accurate quantification and/or localization. To overcome this limitation, we used expert annotations in the form of representative localized bounding boxes to further validate and the gingival inflammation ground truth labels. An increased number of representative bounding boxes, or even full localized expert segmentations, may help in reducing this bias. The low number of images used to train the classifier, combined with the widespread poor oral health of the patient population, resulted in the overfitting of the classifier to the few examples of certain classes (MGI 4). An evenly distributed larger dataset should solve this problem. Similarly, a greater variety in camera angles and positions in the mouth of photographed gingiva in the expanded dataset would most likely increase the robustness.

V. CONCLUSION

The paper proposes an automated system that performs pixel-wise segmentation of dental images and successfully segments gingival inflammation from healthy gums. This is the first description, to our knowledge, of segmenting gingival diseases using oral images. The performance of the classifier was validated using multiple images, demonstrating its accuracy and efficacy to distinguish between inflamed gingiva from health gingiva and works better than random chance. Our automated system can use intraoral images in point-of-care settings for early detection of gingival inflammation in patients and help prevent severe periodontal diseases and ultimately loss of tooth.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Roma Bagi and Keith Angelino for clinical expertise and technical help.

REFERENCES

[1] R. C. Page, "Gingivitis," *Journal of Clinical Periodontology*, vol. 13, no. 5, pp. 345–355, 1986.

[2] T. Eckhard, E. M. Valero, and J. L. Nieves, "Labial teeth and gingiva color image segmentation for gingival health-state assessment," in *Conference on Colour in Graphics, Imaging, and Vision*, vol. 2012, no. 1. Society for Imaging Science and Technology, 2012, pp. 102–107.

[3] P. Rechmann, S. W. Liou, B. M. Rechmann, and J. D. Featherstone, "Performance of a light fluorescence device for the detection of microbial plaque and gingival inflammation," *Clinical oral investigations*, vol. 20, no. 1, pp. 151–159, 2016.

[4] "The good practitioners guide to periodontology," https://www.bsperio.org.uk/publications/good_practitioners_guide_2016.pdf?v=3.

[5] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *Journal of pathology informatics*, vol. 7, 2016.

[6] "Soprocare device," https://www.acteongroup.com/us/my-products/imaging/diagnostic-camera/soprocare.

[7] M. Juan, M. Alcaniz, C. Monserrat, V. Grau, and C. Knoll, "Computer-aided periodontal disease diagnosis using computer vision," *Computerized medical imaging and graphics*, vol. 23, no. 4, pp. 209–217, 1999.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.

[11] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv preprint arXiv:1511.00561*, 2015.

[12] R. Lobene, T. Weatherford, N. Ross, R. Lamm, and L. Menaker, "A modified gingival index for use in clinical trials." *Clinical preventive dentistry*, vol. 8, no. 1, pp. 3–6, 1986.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[14] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV), 2016 Fourth International Conference on*. IEEE, 2016, pp. 565–571.

[15] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.