A Deep Learning Approach to Automatic Gingivitis Screening based Onclassification and Localization in RGB Photos

Wen Li

Nanjing Stomatological Hospital, Medical School of Nanjing University

Yuan Liang

University of California, Los Angeles

Xuan Zhang

Nanjing Stomatological Hospital, Medical School of Nanjing University

Chao Liu

Nanjing Stomatological Hospital, Medical School of Nanjing University

Lei He

University of California, Los Angeles

Leiying Miao (✉ miaoleiying80@163.com)

Nanjing Stomatological Hospital, Medical School of Nanjing University

Weibin Sun

Nanjing Stomatological Hospital, Medical School of Nanjing University

Research Article

Keywords: gingivitis, Routine dental visit, diagnosis , convolutional neural network (CNN) model

Posted Date: December 7th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-117989/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on August 19th, 2021.
See the published version at <https://doi.org/10.1038/s41598-021-96091-3>.

Abstract

Routine dental visit is the most common approach to detect the gingivitis. However, such diagnosis can sometimes be unavailable due to the limited medical resources in certain areas and costly for low-income populations. To increase the availability, this study proposes to screen the existence of gingivitis and its irritants, i.e., dental calculus and soft deposits, from oral photos with a novel Multi-Task Learning convolutional neural network (CNN) model. With 625 patients included in this study, the classification Area Under the Curve (AUC) for detecting gingivitis, dental calculus and soft deposits were 87.11%, 80.11%, and 78.57%, respectively; Meanwhile the box-wise localization sensitivity for gingivitis and dental calculus were 66.57% and 45.61%. Moreover, according to a consistency evaluation with three board-certificated dentists, the model achieved a median score of 3.0/5.0 for reasoning locations of soft deposits without any spatial supervision. By comparing to general-purpose CNNs, we showed our model significantly outperformed on both classification and localization tasks, which indicates the effectiveness of Multi-Task Learning on dental disease detection. The results show the potential of deep learning for enabling cost-effective screening of dental diseases among large populations.

Introduction

Gingivitis is the chronic infection of oral gum that has been affecting the public health worldwide, whose main manifestations are redness, bleeding, and bad breath. The latest epidemiological survey on oral health showed that 87.4% of adults aged 35–44 years suffers from the resultant gingival bleeding¹. The development of gingivitis is a continuous progress, with dental calculus and soft mucinous deposits being the main irritants². Soft deposits is a visible bacterial mass on teeth, while dental calculus is a mineralized soft deposit that further destroys the periodontal tissues by constantly adsorbing calcium compounds from saliva. The awareness of gingivitis and its irritants can be helpful for individuals to stall and control the progress of gingivitis with corresponding interventions, as well as getting timely treatment to avoid tooth loss. Currently, routine dental visit is the most effective way for such detection. However, the dental diagnosis is not always available due to the limited medical resources in some regions. Also, it can add economic burdens for individuals of low-income, which might prevent them from routine dental visits³. Therefore, there is a need for computer-assisted system that is cost-effective and can automatically screen for the gingivitis and its irritants, i.e., dental calculus and soft deposits. Based on the results, such systems can also show targeted health-enhancing activities, proper hygiene routines, and clinical treatments, which will be meaningful for promoting the public dental health.

Our work employs deep learning (DL) algorithms for highly efficient and accurate disease detection⁶. Currently, DL has been widely used to detect health issues from imagery captured with cameras or smartphones for everyday users as a supplement to clinical visits^{7,8}. Indeed, BiliScreen achieved the automatic live disorder detection by capturing Jaundice color changes⁹; Vardell introduced a skin disease diagnosis system with skin photos as a source of input¹⁰; and Mariakakis et. al. developed the

fast traumatic brain injury detection by training a model taken pupil images as input¹¹. Despite those works, the use of DL models for screening oral conditions is still much under-explored.

There also exists continuous efforts for enabling the automatic dental diagnosis with DL algorithms. For example, Joachim Krois¹² applied CNNs to detect periodontal bone loss (PBL) on panoramic dental radiographs, while Casalegno et. al. performed caries segmentation from Near-Infrared-Light Transillumination (NILT) images¹³. Jae-Hong Lee evaluate the efficacy of deep CNN algorithms for detection and diagnosis of dental caries on periapical radiographs¹⁴. Yu et. al. also evaluated the performance of CNNs for the skeletal classification with lateral cephalometry¹⁵. Different from the previous work, our task takes imageries as input, which can be less standard in distribution comparing to the medical imaging.

In this work, we initialize the study of applying DL on oral photos for screening gingivitis, dental calculus, and soft deposits. We formulated the task as mixture of dental condition classification and localization by considering the nature of the conditions, and developed a Multi-Task Learning to solve the two different types of tasks with an integrated model. We proved the effectiveness of our model by comparing with general-purpose CNNs and carrying out ablation tests. With the designed system, we expect to bring up the discussion for integrating deep learning into tools for improving public dental health.

Materials And Methods

Acquisition of Data

A total of 3932 oral photos were captured from 625 patients admitted at Department of Periodontics, orthodontics and endodontics, Nanjing Stomatological Hospital, Nanjing University, between January 2018 and December 2019. All the photos were captured by postgraduate dental students and dentists, and the patient's ages cover a range from 14 to 60. The project was approved by the Ethical Review Board at Nanjing University (approval 2019NL-065(KS)). The methods were conducted in accordance with the approved guidelines, written informed consent was obtained from each participant. For children under the age of 18, the written informed consent was obtained from their parents/guardians. To approximate the image quality in the practical scenario, the photos were collected with various equipments which include iPhone 8, iPhone 7, Samsung Galaxy s8, and Canon 6D. No specific in- or exclusion criteria about images, e.g. lighting and resolution, was applied. Three dental diseases were considered in this study: gingivitis, dental calculus, and soft deposit. Among the dataset, 3,175 photos show gingivitis, 921 show dental calculus, and 746 images show soft deposits. Note that each photo can show none, one or more types of conditions. All photos were pseudonymized and no other image processing steps are performed.

We split the data into training, validation and testing subsets by randomly splitting the patients into three independent groups. All photos of each patient only existed in one of the three subsets. Table 1 shows

the patient and photo split. Table 2 shows the distribution of photos with positive findings among the dataset and different subsets.

Table 1
Numbers of Patients and Images Assigned to Training, Validation, and Testing Subsets.

	Training	Validation	Testing	Total
Patients	344 (55.04%)	94 (15.04%)	187 (29.92)	625 (100%)
Images	2138 (54.37%)	608 (15.46%)	1186 (30.16%)	3932 (100%)

Table 2
Distribution of images with positive findings among training, validation, and testing subsets for each type of diagnosis. The data shows the numbers of images with positive findings, as well as their proportions within all the images of a category.

	Gingivitis	Dental Calculus	Soft Deposit
Training	1726 (80.73%)	514 (24.04%)	424 (19.83)
Validation	469 (77.14%)	175 (28.78%)	116 (19.08%)
Testing	980 (82.63%)	232 (19.56%)	206 (17.37%)
Total	3175 (80.75%)	921 (23.42%)	746 (18.97%)

Ground Truth Annotations.

We collected the reference annotations of the three dental conditions for all the photos from three board-certified dentists. In specific, the dataset was evenly split and assigned to the three dentists. Each image was independently labeled by one of the dentists with the given clinical report using the labeling software LabelBox (Labelbox, Inc, CA). For gingivitis and dental calculus, we collected the annotations of bounding boxes for indicating the localizations of diseases. Note that since there can be no well-defined boundaries for diseases in some cases, we followed a common approach¹⁶ by instructing the dentists to focus on the correctness of box centers. Meanwhile, for soft deposit, we only collected image-level classification labels, since such condition mostly appears all over cavities and its localizations are labor-costly to label.

Problem Formulation and Model Architecture.

We formulated the problem as a mixture of object localization and image classification. In specific, we developed a CNN with Multi-task Learning (MTL)^{17,18} for solving both tasks in order to increase the generalization¹⁹ and compactness of the model. Figure 1 shows the overall architecture of our MTL

model, which takes oral images (Figure 1(a)) as input, and outputs both diagnosis and locations of the detected conditions (Figure 1(e)). The model is consisted of three subnets: (i) FNet (feature extraction subnet), (ii) LNet (localization subnet) and (iii) CNet (classification subnet). FNet (Figure 1(b)) extracts deep features of the input image through a stack of convolutional layers, and was trained to be discriminative for both localization and classification tasks. Meanwhile, LNet (Figure 1(c)) regresses over the feature maps derived from FNet for a set of location vectors, where each vector y encodes one bounding box by its coordinates, height, width, and probability for gingivitis and dental calculus. Similar with²⁰, the proposed bounding boxes are aligned to nearest ground-truth boxes during training to approximate localizations, and are filtered with Non-maximum Suppression (NMS)²¹ during inference to reduce overlapped findings. CNet (Figure 1(d)) performs fully connected operations over the extracted feature maps for a length 1 vector as outputs, whose value represented the probability for the existence of soft deposits.

To help users comprehensively understand the diagnosis results, we aim to highlight the spatial locations of the detected dental conditions. For gingivitis and dental calculus, the bounding boxes from the model can already localize the ROIs. However, for soft deposit, the model only produces classification results since their ground-truth location maps are labor costly to annotate. Thus, gradient-based class activation maps²² was used to reason the areas of the images that are most indicative to the classification.

Figure 1(e) shows an example result from our system. The detected gingivitis and dental calculus are pinpointed with boxes, and soft deposits are hinted with a heat-map, where a higher temperature indicates the stronger relevance of a region.

Implementation and Training Strategy.

To train the model, we defined the loss as an equally weighted sum of smooth L1 loss for bounding box regression²⁰, and cross entropy loss for classification²³. We employed intensive augmentations to input images²⁴, which includes random shifts, crops, rotations, scaling, and color channel shifts (random changes of hue, saturation, and exposure). Such augmentations is targeted to increase the robustness of the model for in-the-wild application. Moreover, we employed transfer learning by initializing our FeatNet from VGG-16²⁵ that pre-trained on large-scale image recognition tasks for speeding up the training process²⁶.

The CNN model was developed using the PyTorch framework. The model was trained using a mini batch size of 16 per GPU on three Nvidia 1080 Ti GPUs. Validation set was used to determine the early stopping of the training process. Parameter updates were calculated using the Adam algorithm, with the learning rate set to 1e-4 and decay rate set to 5e-4.

Evaluation Metrics and Statistical Analysis.

We evaluated the model from two aspects: (i) classification performance for telling the existence of a condition, and (ii) localization performance for indicating regions on images that related to a diagnosis.

In terms of the classification performance, we utilized the Receiver Operating Characteristic (ROC) curve, which shows the true-positive rate (TPR), or sensitivity, against its false-positive rate (FPR), or 1 - specificity, as a function of varying discrimination thresholds. For gingivitis and dental calculus, we took the highest probability of the detected bounding boxes as the predicted probability of classification. AUC was used to compare between models.

In terms of localization performance, we utilized the Free-Response ROC (FROC) curve, which shows the bounding-box-wise TPR, or sensitivity, against the average number of false-positive (FP) boxes per image, as a function of varying thresholds for box probabilities. By following the practice of van Ginneken, a predicted box was taken as a hit if its center falls into the range of a ground-truth box. Similarly, the Area Under FROC (FAUC) was used as the index of performance, which was defined as the average sensitivity at false positive rates of 1/2, 1, 2, and 3. This metric is used as an index to compare different models, where a larger value indicates the better performance. To measure the quality of soft deposit predictions, we followed Selvaraju²² by collecting agreement ratings from three board-certified dentists for each localization heat-map of testing images. Specifically, we show dentists images that were detected with soft deposits together with the localization heat-maps that visualized as in Figure 1(e). Then a rating is given on a scale from 1 (strongly disagree) to 5 (strongly agree) by evaluating if a heat-map demonstrates the regions of the condition according to dentists' opinions.

Results

Table 3 and Table 4 show the classification and localization performance of different models, respectively. Comparing to the general-purpose classification CNNs (VGG-16²¹ and Residual-50²⁷) and localization CNNs (SSD²⁰), our model has the advantage of handling both types of tasks. The model achieved classification AUC (95% CI) of 87.11 (82.27 to 91.49) for gingivitis, 80.11% (CI: 75.99% to 84.45%) for dental calculus, and 78.57% (CI: 74.32% to 82.78%) for soft deposits; meanwhile the model performed at FAUC (95% CI) of 58.19% (56.15% to 60.20%) and 49.39% (44.40% to 54.69%) for localizing gingivitis and dental calculus, respectively. All the scores were highest scores among different methods, suggesting the effectiveness of the proposed system. Additionally, we conducted an ablation test comparing the performance of our model (FeatNet+ClassNet+LocateNet) with its subnets that trained for classification (FeatureNet+classNet) or localization (FeatureNet+locateNet) solely. The results show that our model outperformed in all metrics, with AUC boosts ranging from 1.20% to 8.82%, and FAUC boosts ranging from 1.26% to 5.52%. The results confirm the advantage of handling multiple tasks simultaneously with joint optimization.

Table 3

Summary of Classification Performance of Different Models. Classification AUC (in Percentage), Sensitivity (Sens.), Specificity (Specif.) are measured for Gingivitis (GI), Dental Calculus (DC), and Soft Deposits (SD).

	GI AUC (95% CI) / %	GI Sens.	GI Specif.	DC AUC (95% CI) / %	DC Sens.	DC Specif.	SD AUC (95% CI) / %	SD Sens.	SD Specif.
VGG-16	77.55 (73.97 to 79.05)	0.682	0.620	74.17 (72.71 to 76.75)	0.645	0.572	72.07 (69.76 to 76.34)	0.679	0.605
Residual-50	83.80 (80.69 to 86.37)	0.755	0.628	78.22 (75.91 to 82.02)	0.780	0.567	71.49 (67.86 to 75.98)	0.641	0.586
FNet+CNet	84.28 (80.94 to 87.11)	0.802	0.578	76.78 (74.50 to 79.47)	0.745	0.615	69.75 (68.29 to 71.57)	0.740	0.480
Ours (high-sensitivity operation point)	87.11 (82.27 to 91.49)	0.878	0.639	80.11 (75.99 to 84.45)	0.778	0.655	78.57 (74.32 to 82.78)	0.787	0.590
Ours (high-specifity operation point)	87.11 (82.27 to 91.49)	0.601	0.839	80.11 (75.99 to 84.45)	0.542	0.836	78.57 (74.32 to 82.78)	0.565	0.800

Table 4

Summary of Localization Performance of Different Models. Detection FAUC (in Percentage), Bounding-box-wise Sensitivity (Sens.), Average False Positives (Avg. False Positives) are measured for Gingivitis (GI) and Dental Calculus (DC).

	GI FAUC (95% CI) / %	GI Sens.	GI Avg. False Positives	DC FAUC (95% CI) / %	DC Sens.	DC Avg. False Positives
SSD	44.19 (41.19 to 47.08)	0.422	1.280	36.23 (32.93 to 37.89)	0.375	1.721
FNet+LNet	56.93 (56.05 to 59.42)	0.565	1.340	43.87 (41.26 to 48.33)	0.392	1.150
Ours (high-sensitivity operation point)	58.19 (56.15 to 60.20)	0.666	1.522	49.39 (44.40 to 54.69)	0.456	1.330
Ours (high-specificity operation point)	58.19 (56.15 to 60.20)	0.432	0.585	49.39 (44.40 to 54.69)	0.380	0.260

The ROC and FROC curves for each diagnosis are detailed in Figure 2(a) and Figure 2(b), with the selected operating points of shown with black diamonds and red dots. Two types of operating points were designed by following^{28,29}: the high-specificity operating point with a higher discrimination threshold that aims for reducing false positives, and the high-sensitivity operating point with a lower discrimination threshold for keeping the missing rate low. The model achieved the mean specificities of 83.87%, 83.61%, and 79.98% under the high-specificity operating points, meanwhile the mean sensitivities of 87.83%, 77.79%, and 78.68% under the high-sensitivity operating points, both for gingivitis, dental calculus, and soft deposit, respectively. For localization performance, the model achieved the mean box-wise sensitivities of 66.57% and 45.61% for gingivitis and dental calculus, respectively, at the high-sensitivity operating point. Moreover, Figure 2(c) shows the dentists' ratings on the attention-based localization for soft deposits. While the attention-based method has been widely applied to interpret CNNs^{30,31}, it lacks formal evaluations of location-indicating accuracy for the dental diagnosis purpose. According to the experiment, our model achieved scores with a median of 3.00, mean of 2.81, and std of 1.02, on a scale from 1 to 5. Based on the scorers' feedback, the following two factors can lead to the lower localization scores. First, different from bounding boxes that pinpointing the exact locations, the heat-maps can only circle out areas with larger ranges. Second, the model attention can often localize only part of the related regions. This can be explained as the model does not count on all regions for reaching a classification results^{22,30}.

Figure 3 depicts selected results obtained on the testing images for a qualitative overview of our model's performance. Ground-truth annotations, heat-map predictions and bounding box predictions are shown in the left, middle, and right column. By looking into the outputs, we found that the over-exposure, under-exposure and incorrect focus of photos can lead to wrong predictions.

Discussion And Future Work

Previous studies have explored predicting gum health with self-reported questionnaires^{4,5}. Their results have shown that several self-reported measures and risk factors are strongly related to the presence of gum diseases. Different from those works, we aim to predict gingivitis as well as its irritants as early indicators from oral photos by learning their common appearance patterns. The method can be promising, since the smartphones have become increasingly low-cost and ubiquitous recently. Moreover, the algorithm can also be complementary with the traditional questionnaire-based detections for higher reliability and accuracy.

Our work targets at a convenient and efficient dental condition screening tool that can be used by the public. As such, showing localization of detected conditions is particularly important, because it can help users without medical background better understand the screening results. Moreover, it has been shown that the better explainability of model's outputs can help gain trust of users to a system³². We formulated the localization of gingivitis and dental calculus as bounding box regression by considering the appearance of the conditions and saving labour cost. For soft deposits, we formulated the task as image-wise classification, and reasoned its locations with model attentions.

To improve the system efficiency, we employed Multi-Task Learning, such that both types of tasks, i.e. classification and localization, can be solved with one integrated CNN model. Moreover, our experiments indicated that our model also outperformed the state-of-the-art CNNs that carried out single type of task in accuracy, mainly because the co-optimization of multiple tasks increases the model generalization. We further confirmed this with ablation tests, where the model with MTL showed significant accuracy improvements comparing to its subnets that trained for classification or localization solely. We believe the findings can help with the CNN design for other dental diagnosis with multiple goals.

Our work still exhibits several shortcomings, and we discuss the possible solutions for future researches. First, our dataset is limited in the sense that the photos were collected from a single organization, and currently it only covered age range from 14 to 60. We have the plan to further enrich the dataset for a wider age range from multiple sites globally. Second, our model achieved a relative low accuracy on soft deposit for the localization task, partially due to the lack of spatial annotations as the guidance for training supervision. Instead of collecting pixel-wise segmentation maps, which can be extremely labour costly, we advocate that future studies could apply recently proposed weakly-supervised learning to train with low-quality spatial annotations, e.g. partial labels over images to indicate several typical areas of diseases³³. Moreover, the model could also benefit from semi-supervised learning by augmenting a part of dataset with pixel-wise labels^{34,35}, while the other part only comes with image-wise labels. Third, the

current model cannot utilize data modality other than images for diagnosis. Encoding³⁶ and fusing of medical history and self-reported symptoms of a patient into CNNs could be promising to improve the model accuracy³⁷.

Conclusion

In this study, a deep learning model for the detection of gingivitis, dental calculus, and soft deposits from oral photos was proposed. We formulated the model with Multi-Task Learning, which effectively improve its compactness and accuracy. We evaluated our model for both classification and localization tasks. Based on the results, we show deep learning is promising for enabling the cost-effective screening of dental diseases among large populations, particularly with continual improvements of data quantity and model architectures.

Declarations

Data Availability

The data used in current study were collected from Medical School of Nanjing University and is available only for the granted research. However, the data can be made available if requested within data protection and regulation guideline.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (51772144); Project of Invigorating Health Care through Science, Technology and Education Jiangsu Provincial Medical Youth Talent (QNRC2016120). The authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article.

Author contributions

Wen Li and Yuan Liang, prepared the data, conducted the modeling, and analyzed the results, Xuan Zhang and Chao Liu, contributed to data acquisition, critically revised the manuscript; Lei He contributed to data acquisition and critically revised the manuscript; Weibin Sun and Leiying Miao, drafted and provided expert clinical guidance to the study. All authors gave final approval and agree to be accountable for all aspects of the work and ensure that questions related to the accuracy or integrity of any part of the work.

Competing interests

The authors declare no competing interests.

References

1. Supranoto, S.C., Slot, D.E., Addy, M. & Ga, V.D.W. The effect of chlorhexidine dentifrice or gel versus chlorhexidine mouthwash on plaque, gingivitis, bleeding and tooth discoloration: a systematic review. *International Journal of Dental Hygiene*. **13**, 83–92 (2015).
2. Miyauchi, S. et al. Sphingomyelin phosphodiesterase 3 enhances cytodifferentiation of periodontal ligament cells. *J Dent Res*. **96**, 339–346 (2017).
3. Petersen, P. E., Bourgeois, D., Ogawa, H., Estupinan-Day, S. & Ndiaye, C. The global burden of oral diseases and risks to oral health. *Bulletin of the World Health Organization*. **83**, 661–669 (2005).
4. Beltrán-Aguilar, E.D., Eke, P.I., Thornton-Evans, G. & Petersen, P. E. Recording and surveillance systems for periodontal diseases. *Periodontol 2000*. **60**, 40–53 (2012).
5. Eke, P.I. et al. Self-reported measures for surveillance of periodontitis. *J Dent Res*. **92**, 1041–1047 (2013).
6. Liu, Y.P., Li, Z., Xu, C., Li, J. & Liang, R. et al. Referable diabetic retinopathy identification from eye fundus images with weighted path for convolutional neural network. *Artificial Intelligence in Medicine*. **99**, 101694 (2019).
7. Litjens, G. et al. A survey on deep learning in medical image analysis. *Med Image Ana*. **42**, 60–88 (2017).
8. Wang, G. et al. Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE Trans Med Imaging*. **37**, 1562–1573 (2018).
9. Mariakakis, A. et al. Biliscreen: smartphone-based scleral jaundice monitoring for liver and pancreatic disorders. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, **1**, 20 (2017).
10. Vardell, E. & Bou-Crick, C. VisualDx: a visual diagnostic decision support tool. *Medical reference services*; **31**, 414–424 (2012).
11. Mariakakis, A. et al. PupilScreen: Using Smartphones to Assess Traumatic Brain Injury. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. **1**, 81 (2017).
12. Krois, J. et al. Deep learning for the radiographic detection of periodontal bone loss. *Sci Rep*. **9**, 84–95 (2019).
13. Casalegno, F. et al. Caries detection with near-infrared transillumination using deep learning. *J Dent Res*. **98**, 1227–1233 (2019).
14. Lee, J.H., Kim, D.H., Jeong, S.N. & Choi, S.H. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J Periodontal Implant Sci*. **48**, 114–123 (2018).
15. Yu, H.J. et al. Automated skeletal classification with lateral cephalometry based on artificial intelligence. *J Dent Res*. **99**, 249–256 (2020).

16. Armato, S.G. et al. The lung image database consortium (LIDC): Ensuring the integrity of expert-defined "truth"¹. *Academic radiology*. **14**, 1455–1463 (2007).
17. Ranjan, R., Sankar, S., Castillo, C. D. & Chellappa, R. An all-in-one convolutional neural network for face analysis. *2017 IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 17–24 (IEEE, 2017).
18. Bansal, A., Nanduri, A., Castillo, C., Ranjan, R. & Chellappa, R. Umdfaces: An annotated face dataset for training deep networks. *2016 IEEE International Joint Conference on Biometrics (IJCB)*, (IEEE, 2016).
19. Krizhevsky, A., Sutskever, I., & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*. **25**, 1097–1105 (2012).
20. Liu, W., Anguelov, D., Erhan, D., Szegedy, C. & Berg, A. C. SSD: Single shot multibox detector. *In European conference on computer vision*, 21–37 (Springer, 2016).
21. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large scale image recognition. *Computer Science. arXiv preprint arXiv:1409.1556* (2014).
22. Selvaraju, R.R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE international conference on computer vision*, 42–48 (IEEE, 2017).
23. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. **542**, 115–118 (2017).
24. Redmon, J. & Farhadi, A. YOLO9000: better, faster, stronger. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 7263–7271 (IEEE, 2017).
25. Deng, J. et al. Imagenet: a large scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 643–651 (IEEE, 2009).
26. van Ginneken, B. et al. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. *Medical image analysis*. **14**, 707–722 (2010).
27. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (IEEE, 2016).
28. Gulshan, V. et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*. **316**, 2402–2410 (2016).
29. De Vrijer, M., Medendorp, W. P. & Van Gisbergen, J. A. M. Accuracy-precision trade-off in visual orientation constancy. *Journal of vision*. **9**, 9–9 (2009).
30. Montavon, G., Samek, W. & Müller, K. R. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*. **73**, 1–15 (2018).
31. Jalali, A., et al. Deep learning for improved risk prediction in surgical outcomes. *Sci Rep*. **10**, 9289 (2020).

32. Gilpin, L.H. et al. Explaining explanations: an overview of interpretability of machine learning. *Computer Science and Artificial Intelligence Laboratory Massachusetts Institute of Technology Cambridge*, (2018).
33. Khoreva, A., Benenson, R., Hosang, J., Hein, M. & Schiele, B. Simple does it: Weakly supervised instance and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 825–833 (IEEE, 2016).
34. Bai, W. et al. Semi-supervised learning for network-based cardiac MR image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 253–260 (Springer, 2017).
35. Cheplygina, V., de Bruijne, M. & Pluim, J.P.W. Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical image analysis*. **54**, 280–296 (2019).
36. Choi, E., Xiao, C., Stewart, W.F. & Sun, J. Mime: Multilevel medical embedding of electronic health records for predictive healthcare. *Advances in neural information processing systems*. **4**, 456–461 (2018).
37. Esteva, A. et al. A guide to deep learning in healthcare. *Nat Med*. **25**, 24–29 (2019).

Figures

Figure 1

Overview of the model architecture. The model consists of three subnets: (b) FNet of feature extraction, (c) LNet for bounding box-based localization, and (d) CNet for classifying the existence of conditions. Given an (a) oral cavity image as input, (e) the model outputs both the probabilities for diagnosis, as well as the locations of the detections with heat-maps and bounding boxes. The numbers at model building blocks mark for channel dimensions. Gingivitis, dental calculus, and soft deposits are represented as GI, CA, and SD, respectively.

Figure 2

Quantitative analysis of the results of our model. (a) Receiver operating characteristic curves for predicting the existence of conditions. (b) Free-Response receiver operating characteristic curves for localizing conditions. (c) Distribution of dentists' ratings on the localization for soft deposits. Gingivitis, dental calculus, and soft deposits are represented as GI, DC, and SD, Respectively.

Figure 3

Examples of detection results on the testing data. Left: ground-truth annotations. Middle: Predicted locations of soft deposits represented with heat-map. Right: predicted locations of gingivitis and dental calculus represented with green and blue boxes, respectively.