HC-Net: Hybrid Classification Network for Automatic Periodontal Disease Diagnosis

Lanzhuju Mei[1,2,4], Yu Fang[1,2,4], Zhiming Cui[1], Ke Deng[3], Nizhuan Wang[1], Xuming He[2], Yiqiang Zhan[4], Xiang Zhou[4], Maurizio Tonetti[3(✉)], and Dinggang Shen[1,4,5(✉)]

[1] School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
dgshen@shanghaitech.edu.cn
[2] School of Information Science and Technology, ShanghaiTech University, Shanghai, China
[3] Shanghai Ninth People's Hospital, Shanghai Jiao Tong University, Shanghai, China
maurizio.tonetti@ergoperio.eu
[4] Shanghai United Imaging Intelligence Co. Ltd., Shanghai, China
[5] Shanghai Clinical Research and Trial Center, Shanghai, China

Abstract. Accurate periodontal disease classification from panoramic X-ray images is of great significance for efficient clinical diagnosis and treatment. It has been a challenging task due to the subtle evidence in radiography. Recent methods attempt to estimate bone loss on these images to classify periodontal diseases, relying on the radiographic manual annotations to supervise segmentation or keypoint detection. However, these radiographic annotations are inconsistent with the clinical golden standard of probing measurements and thus can lead to measurement errors and unstable classifications. In this paper, we propose a novel hybrid classification framework, HC-Net, for accurate periodontal disease classification from X-ray images, which consists of three components, i.e., tooth-level classification, patient-level classification, and a learnable adaptive noisy-OR gate. Specifically, in the tooth-level classification, we first introduce instance segmentation to capture each tooth, and then classify the periodontal disease in the tooth level. As for the patient level, we exploit a multi-task strategy to jointly learn patient-level classification and classification activation map (CAM) that reflects the confidence of local lesion areas upon the panoramic X-ray image. Eventually, the adaptive noisy-OR gate obtains a hybrid classification by integrating predictions from both levels. Extensive experiments on the dataset collected from real-world clinics demonstrate that our proposed HC-Net achieves state-of-the-art performance in periodontal disease classification and shows great application potential. Our code is available at https://github.com/ShanghaiTech-IMPACT/Periodental_Disease.

1 Introduction

Periodontal disease is a set of inflammatory gum infections damaging the soft tissues in the oral cavity, and one of the most common issues for oral health [4].

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14225, pp. 54–63, 2023.
https://doi.org/10.1007/978-3-031-43987-2_6

If not diagnosed and treated promptly, it can develop into irreversible loss of the bone and tissue that support the teeth, eventually causing tooth loosening or even falling out. Thus, it is of great significance to accurately classify periodontal disease in an early stage. However, in clinics, dentists have to measure the clinical attachment loss (CAL) of each tooth by manual probing, and eventually determine the severity and progression of periodontal disease mainly based on the most severe area [14]. This is excessively time-consuming, laborious, and over-dependent on the clinical experience of experts. Therefore, it is essential to develop an efficient automatic method for accurate periodontal disease diagnosis from radiography, i.e., panoramic X-ray images.

With the development of computer techniques, computer-aided diagnosis has been widely applied for lesion detection [8,9] and pathological classification [10] in medical image analysis. However, periodontal disease diagnosis from panoramic X-ray images is a very challenging task. While reliable diagnosis can only be provided from 3D probing measurements of each tooth (i.e. clinical golden standard), evidence is highly subtle to be recognized from radiographic images. Panoramic X-ray images make it even more difficult with only 2D information, along with severe tooth occlusion and distortion. Moreover, due to this reason, it is extremely hard to provide confident and consistent radiographic annotations on these images, even for the most experienced experts. Many researchers have already attempted to directly measure radiographic bone loss from panoramic X-ray images for periodontal disease diagnosis. Chang et al. [2] employ a multi-task framework to simulate clinical probing, by detecting bone level, cementoenamel junction (CEJ) level, and tooth long axis. Jiang et al. [7] propose a two-stage network to calculate radiographic bone loss with tooth segmentation and keypoint object detection. Although these methods provide feasible strategies, they still rely heavily on radiographic annotations that are actually not convincing. These manually-labeled landmarks are hard to accurately delineate and usually inconsistent with clinical diagnosis by probing measurements. For this reason, the post-estimated radiographic bone loss is easily affected by prediction errors and noises, which can lead to incorrect and unstable diagnosis.

To address the aforementioned challenges and limitations of previous methods, we propose HC-Net, a novel hybrid classification framework for automatic periodontal disease diagnosis from panoramic X-ray images, which significantly learns from clinical probing measurements instead of any radiographic manual annotations. The framework learns upon both tooth-level and patient-level with three major components, including tooth-level classification, patient-level classification, and an adaptive noisy-OR gate. Specifically, tooth-level classification first applies tooth instance segmentation, then extracts features from each tooth and predicts a tooth-wise score. Meanwhile, patient-level classification provides patient-wise prediction with a multi-task strategy, simultaneously learning a classification activation map (CAM) to show the confidence of local lesion areas upon the panoramic X-ray image. Most importantly, a learnable adaptive noisy-OR gate is designed to integrate information from both levels, with the tooth-level

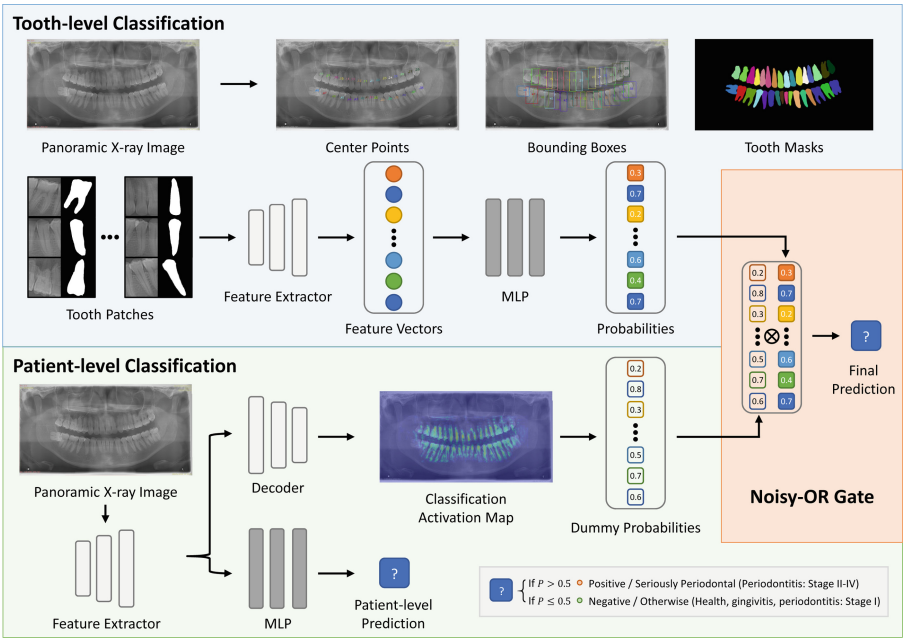

Fig. 1. Illustration of our HC-Net.

scores and patient-level CAM. Note that our classification is only supervised by the clinical golden standard, i.e., probing measurements. We provide comprehensive learning and integration on both tooth-level and patient-level classification, eventually contributing to confident and stable diagnosis. Our proposed HC-Net is validated on the dataset from real-world clinics. Experiments have demonstrated the outstanding performance of our hybrid structure for periodontal disease diagnosis compared to state-of-the-art methods.

2 Method

An overview of our proposed framework, HC-Net, is shown in Fig. 1. We first formulate our task (Sect. 2.1), and then elaborate the details of tooth-level classification (Sect. 2.2), patient-level classification (Sect. 2.3), and adaptive noisy-OR gate (Sect. 2.4), respectively.

2.1 Task Formulation and Method Overview

In this paper, we aim to classify each patient into seriously periodontal (including periodontitis stage II-IV) or not (including health, gingivitis, and periodontitis stage I), abbreviated below as 'positive' or 'negative'. We collect a set of panoramic X-ray images $\mathcal{X} = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_N\}$ with their patient-level

labels $\mathcal{Y} = \{\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_N\}$ from clinical diagnosis, where $\mathcal{Y}_i \in \{0, 1\}$ indicates whether the i-th patient is negative (0) or positive (1). For the i-th patient, we acquire corresponding tooth-level labels $\mathcal{T}_i = \{\mathcal{T}_i^1, \mathcal{T}_i^2, \ldots, \mathcal{T}_i^{K_i}\}$ from clinical golden standard, where K_i denotes the number of teeth, and $\mathcal{T}_i^j \in \{0, 1\}$ indicates whether the j-th tooth of the i-th patient is positive or negative.

Briefly, our goal is to build a learning-based framework to predict the probability $\mathcal{P}_i \in [0, 1]$ of the i-th patient from panoramic X-ray image. An intuitive solution is to directly perform patient-level classification upon panoramic X-ray images. However, it fails to achieve stable and satisfying results (See Sect. 3.2), mainly for the following two reasons. Firstly, evidence is subtle to be recognized in the large-scale panoramic X-ray image (notice that clinical diagnosis relies on tedious probing around each tooth). Secondly, as we supervise the classification with clinical golden standard (i.e., probing measurements), a mapping should be well designed and established from radiography to this standard, since the extracted discriminative features based on radiography may not be well consistent with the golden standard. Therefore, as shown in Fig. 1, we propose a novel hybrid classification framework to learn upon both tooth-level and patient-level, and a learnable adaptive noisy-OR gate that integrates the predictions from both labels and returns the final classification (i.e., positive or negative).

2.2 Tooth-Level Classification

Given the panoramic X-ray image of the i-th patient, we propose a two-stage structure for tooth-level classification, which first captures each tooth with tooth instance segmentation, and then predicts the classification of each tooth. Tooth instance segmentation aims to efficiently detect each tooth with its centroid, bounding box, and mask, which are later used to enhance tooth-level learning. It introduces a detection network with Hourglass [12] as the backbone, followed by three branches, including tooth center regression, bounding box regression, and tooth semantic segmentation. Specifically, the first branch generates tooth center heatmap $\mathcal{H}$. We obtain the filtered heatmap $\tilde{\mathcal{H}}$ to get center points for each tooth, by a kernel that retains the peak value for every 8-adjacent, described as

$$\mathcal{H}_{p_c} = \begin{cases} \mathcal{H}_{p_c}, \text{ if } \mathcal{H}_{p_c} \geq \mathcal{H}_{\mathbf{p}_j}, \forall \mathbf{p}_j \in \mathbf{p} \\ 0, \text{ otherwise} \end{cases}, \tag{1}$$

where we denote $\mathbf{p} = \{p_c + e_i\}_{i=1}^8$ as the set of 8-adjacent, where p_c is the center point and $\{e_i\}_{i=1}^8$ is the set of direction vectors. The second branch then uses the center points and image features generated by the backbone to regress the bounding box offsets. The third branch utilizes each bounding box to crop the original panoramic X-ray image and segment each tooth. Eventually, with the image patch and corresponding mask $\mathcal{A}_i^j$ for the j-th tooth of the i-th patient, we employ a classification network (i.e., feature extractor and MLP) to predict the probability $\tilde{\mathcal{T}}_i^j$, if the tooth being positive. To train the tooth-level framework, we design a multi-term objective function to supervise the learning process. Specifically, for tooth center regression, we employ the focal loss of [16] to calculate

the heatmap error, denoted as $\mathcal{L}_{ctr}$. For bounding box regression, we utilize L1 loss to calculate the regression error, denoted as $\mathcal{L}_{bbx}$. For tooth semantic segmentation, we jointly compute the cross-entropy loss and dice loss, denoted as $\mathcal{L}_{seg} = 0.5 \times (\mathcal{L}_{seg_{CE}} + \mathcal{L}_{seg_{Dice}})$. We finally supervise the tooth-level classification with a cross-entropy loss, denoted as $\mathcal{L}_{cls_t}$. Therefore, the total loss of the tooth-level classification is formulated as $\mathcal{L}_{tooth} = \mathcal{L}_{ctr} + 0.1 \times \mathcal{L}_{bbx} + \mathcal{L}_{seg} + \mathcal{L}_{cls_t}$.

2.3 Patient-Level Classification

As described in Sect. 2.1, although patient-level diagnosis is our final goal, direct classification is not a satisfying solution, and thus we propose a hybrid classification network on both tooth-level and patient-level. Additionally, to enhance patient-level classification, we introduce a multi-task strategy that simultaneously predicts the patient-level classification and a classification activation map (CAM). The patient-level framework first utilizes a backbone network to extract image features for its following two branches. One branch directly determines whether the patient is positive or negative through an MLP, which makes the extracted image features more discriminative. We mainly rely on the other branch, which transforms the image features into CAM to provide local confidence upon the panoramic X-ray image.

Specifically, for the i-th patient, with the predicted area $\{\mathcal{A}_i^j\}_{j=1}^{K_i}$ of each tooth and the CAM $\mathcal{M}_i$, the intensity $\mathcal{I}$ of the j-th tooth can be obtained, described as

$$\mathcal{I}_i^j = \mathcal{C}(\mathcal{M}_i, \mathcal{A}_i^j), \tag{2}$$

where $\mathcal{C}(\cdot, *)$ denotes the operation that crops $\cdot$ with the area of $*$. To supervise the CAM, we generate a distance map upon the panoramic X-ray image, based on Euclidean Distance Transform with areas of positive tooth masks. In this way, we train patient-level classification in a multi-task scheme, jointly with direct classification and CAM regression, which increases the focus on possible local areas of lesions and contributes to accurate classification. We introduce two terms to train the patient-level framework, including a cross-entropy loss $\mathcal{L}_{cls_p}$ to supervise the classification, and a mean squared loss $\mathcal{L}_{CAM}$ to supervise the regression for CAM. Eventually, the total loss of the patient-level classification is $\mathcal{L}_{patient} = \mathcal{L}_{cls_p} + \mathcal{L}_{CAM}$.

2.4 Learnable Adaptive Noisy-OR Gate

We finally present a learnable adaptive noisy-OR gate [13] to integrate tooth-level classification and patient-level classification. To further specify the confidence of local lesion areas on CAM, we propose to learn dummy probabilities $\mathcal{D}_i^j$ for each tooth with its intensity $\mathcal{I}_i^j$

$$\mathcal{D}_i^j = \Phi(\mathcal{I}_i^j), \tag{3}$$

where Φ denotes the pooling operation.

In this way, as shown in Fig. 1, we have obtained tooth-wise probabilities predicted from both tooth-level (i.e., probabilities $\{\tilde{\mathcal{T}}_i^j\}_{j=1}^{K_i}$) and patient-level (i.e., dummy probabilities $\{\mathcal{D}_i^j\}_{j=1}^{K_i}$). We then formulate the final diagnosis as hybrid classification, by designing a novel learnable adaptive noisy-OR gate to aggregate these probabilities, described as

$$\tilde{\mathcal{Y}}_i = 1 - \prod_{j \in \mathcal{G}_i} \mathcal{D}_i^j (1 - \tilde{\mathcal{T}}_i^j), \tag{4}$$

where $\tilde{\mathcal{Y}}_i$ is the final prediction of the i-th patient, $\mathcal{G}_i$ is the subset of tooth numbers. We employ the binary cross entropy loss $\mathcal{L}_{gate}$ to supervise the learning of adaptive noisy-OR Gate. Eventually, the total loss $\mathcal{L}$ of our complete hybrid classification framework is formulated as $\mathcal{L} = \mathcal{L}_{tooth} + \mathcal{L}_{patient} + \mathcal{L}_{gate}$.

3 Experiments

3.1 Dataset and Evaluation Metrics

To evaluate our framework, we collect 426 panoramic X-ray images of different patients from real-world clinics, with the same size of 2903 × 1536. Each patient has corresponding clinical records of golden standard, measured and diagnosed by experienced experts. We randomly split these 426 scans into three sets, including 300 for training, 45 for validation, and 81 for testing. To quantitatively evaluate the classification performance of our method, we report the following metrics, including accuracy, F1 score, and AUROC. Accuracy directly reflects the performance of classification. F1 score further supports the accuracy with the harmonic mean of precision and recall. AUROC additionally summarizes the performance over all possible classification thresholds.

3.2 Comparison with Other Methods

We mainly compare our proposed HC-Net with several state-of-the-art classification networks, which can be adapted for periodontal disease diagnosis. ResNet [5], DenseNet [6], and vision transformer [3] are three of the most representative classification methods, which are used to perform patient-level classification as competing methods. We implement TC-Net as an approach for tooth-level classification, which extracts features respectively from all tooth patches, and all features are concatenated together to directly predict the diagnosis. Moreover, we notice the impressive performance of the multi-task strategy in medical imaging classification tasks [15], and thus adopt MTL [1] to perform multi-task learning scheme. Note that we do not include [2,7] in our comparisons, as they do not consider the supervision by golden standard and heavily rely on unconvincing radiographic manual annotations, which actually cannot be applied in clinics. We employed the well-studied CenterNet [16] for tooth instance segmentation, achieving promising detection (mAP50 of 93%) and segmentation (DICE of 91%) accuracy.

Table 1. Quantitative comparison with representative classification methods for periodontal disease classification.

Method	Accuracy (%)	F1 Score (%)	AUROC (%)
ResNet [5]	80.25	83.33	91.24
DenseNet [6]	86.42	87.36	93.30
TC-Net	85.19	87.50	93.49
xViTCOS [11]	86.42	88.17	94.24
MTL [1]	87.65	89.80	95.12
HC-Net	**92.59**	**93.61**	**95.81**

As shown in Table 1, our HC-Net outperforms all other methods by a large margin. Compared to the patient-level classification methods (such as, ResNet [5], DenseNet [6] and transformer-based xViTCOS [11]) and the tooth-level classification method (TC-Net), MTL [1] achieves better performance and robustness in terms of all metrics, showing the significance of learning from both levels with multi-task strategy. Compared to MTL, we exploit the multi-task strategy with CAM in the patient-level, and design an effective adaptive noisy-OR gate to integrate both levels. Although the DeLong test doesn't show a significant difference, the boosting of all metrics (e.g., accuracy increase from 87.65% to 92.59%) demonstrates the contributions of our better designs that can aggregate both levels more effectively.

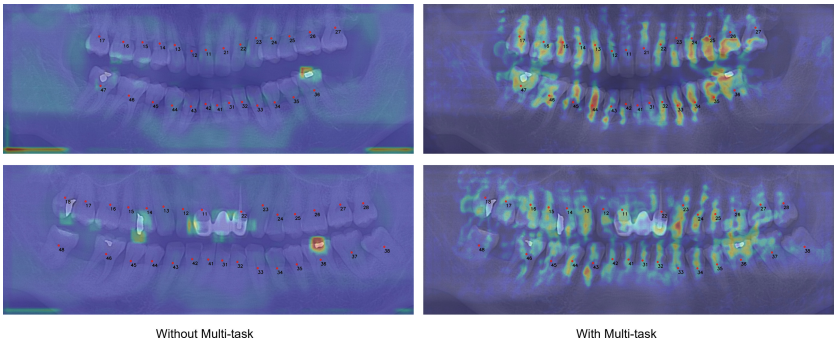

Fig. 2. Illustration of classification activation maps to validate the effectiveness of multi-task strategy with CAM. The first and second columns are visualized respectively from B-Net and M-Net.

3.3 Ablation Studies

We conduct ablative experiments to validate the effectiveness of each module in HC-Net, including patient-level multi-task strategy with classification activation map (CAM) and hybrid classification with adaptive noisy-OR gate. We first define the baseline network, called B-Net, with only patient-level classification. Then, we enhance B-Net with the multi-task strategy, denoted as M-Net, which involves CAM for joint learning on the patient level. Eventually, we extend B-Net to our full framework HC-Net, introducing the tooth-level classification and the adaptive noisy-OR gate.

Effectiveness of Multi-task Strategy with CAM. We mainly compare M-Net to B-Net to validate the multi-task strategy with CAM. We show the classification activation area of both methods as the qualitative results in Fig. 2. Obviously, the activation area of B-Net is almost evenly distributed, while M-Net concentrates more on the tooth area. It shows great potential in locating evidence on local areas of the large-scale panoramic X-ray image, which discriminates the features to support classification. Eventually, it contributes to more accurate qualitative results, as shown in Table 2 and Fig. 3.

Table 2. Quantitative comparison for ablation study.

Method	Accuracy (%)	F1 Score (%)	AUROC (%)
B-Net	86.42	87.36	93.30
M-Net	90.12	91.67	95.37
HC-Net	92.59	93.61	95.81

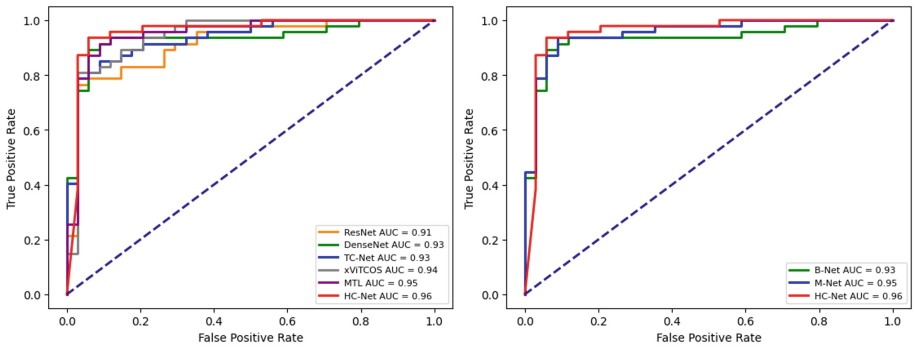

Fig. 3. Illustration of the ROC curves for the comparison and ablation.

Effectiveness of Hybrid Classification with Noisy-OR Gate. We eventually utilize hybrid classification with adaptive noisy-OR Gate, comparing our full framework HC-Net to M-Net. In Table 2 and Fig. 3, we observe that all metrics are dramatically improved. Specifically, the accuracy and F1 score are boosted from 90.12% and 91.67%, to 92.59% and 93.61%, respectively. Note that the AUROC is also significantly increased to 95.81%, which verifies that hybrid classification with noisy-OR gate can improve both the accuracy and robustness of our framework.

3.4 Implementation Details

Our framework is implemented based on the PyTorch platform and is trained with a total of 200 epochs on the NVIDIA A100 GPU with 80GB memory. The feature extractors are based on DenseNet [6]. We use the Adam optimizer with the initial learning rate of 0.001, which is divided by 10 every 50 epochs. Note that in the learnable noisy-or gate, we utilize the probabilities of the top 3 teeth to make predictions for the final outcome.

4 Conclusion

We propose a hybrid classification network, HC-Net, for automatic periodontal disease diagnosis from panoramic X-ray images. In tooth-level, we introduce instance segmentation to help extract features for tooth-level classification. In patient-level, we adopt the multi-task strategy that jointly learns the patient-level classification and CAM. Eventually, a novel learnable adaptable noisy-OR gate integrates both levels to return the final diagnosis. Notice that we significantly utilize the clinical golden standard instead of unconvincing radiographic annotations. Extensive experiments have demonstrated the effectiveness of our proposed HC-Net, indicating the potential to be applied in real-world clinics.

References

1. Sainz de Cea, M.V., Diedrich, K., Bakalo, R., Ness, L., Richmond, D.: Multi-task learning for detection and classification of cancer in screening mammography. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12266, pp. 241–250. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59725-2_24
2. Chang, H.J., et al.: Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis. Sci. Rep. **10**(1), 1–8 (2020)
3. Dosovitskiy, A., et al.: An image is worth 16×16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
4. Eke, P.I., Dye, B.A., Wei, L., Thornton-Evans, G.O., Genco, R.J.: Prevalence of periodontitis in adults in the united states: 2009 and 2010. J. Dent. Res. **91**(10), 914–920 (2012)
5. 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, pp. 770–778 (2016)

6. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
7. Jiang, L., Chen, D., Cao, Z., Wu, F., Zhu, H., Zhu, F.: A two-stage deep learning architecture for radiographic assessment of periodontal bone loss (2021)
8. Kim, J., Lee, H.S., Song, I.S., Jung, K.H.: DeNTNet: deep neural transfer network for the detection of periodontal bone loss using panoramic dental radiographs. Sci. Rep. **9**(1), 1–9 (2019)
9. Krois, J., et al.: Deep learning for the radiographic detection of periodontal bone loss. Sci. Rep. **9**(1), 1–6 (2019)
10. Madabhushi, A., Lee, G.: Image analysis and machine learning in digital pathology: challenges and opportunities. Med. Image Anal. **33**, 170–175 (2016)
11. Mondal, A.K., Bhattacharjee, A., Singla, P., Prathosh, A.: xViTCOS: explainable vision transformer based COVID-19 screening using radiography. IEEE J. Trans. Eng. Health Med. **10**, 1–10 (2021)
12. Newell, A., Yang, K., Deng, J.: Stacked hourglass networks for human pose estimation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9912, pp. 483–499. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_29
13. Srinivas, S.: A generalization of the noisy-or model. In: Uncertainty in Artificial Intelligence, pp. 208–215. Elsevier (1993)
14. Tonetti, M.S., Greenwell, H., Kornman, K.S.: Staging and grading of periodontitis: framework and proposal of a new classification and case definition. J. Periodontol. **89**, S159–S172 (2018)
15. Zhao, Y., Wang, X., Che, T., Bao, G., Li, S.: Multi-task deep learning for medical image computing and analysis: a review. Comput. Biol. Med. **153**, 106496 (2022)
16. Zhou, X., Wang, D., Krähenbühl, P.: Objects as points. arXiv preprint arXiv:1904.07850 (2019)