Received: 6 October 2019 | Revised: 16 November 2019 | Accepted: 22 November 2019

DOI: 10.1111/jop.12983

ORIGINAL ARTICLE

Journal of Oral Pathology & Medicine WILEY

A personalized computational model predicts cancer risk level of oral potentially malignant disorders and its web application for promotion of non-invasive screening

Xiangjian Wang[1,2] | Jin Yang[1] | Changlei Wei[1,3] | Gang Zhou[4] | Lanyan Wu[5] | Qinghong Gao[6] | Xin He[1] | Jiahong Shi[1] | Yingying Mei[1] | Ying Liu[7] | Xueke Shi[1] | Fanglong Wu[1] | Jingjing Luo[1] | Yiqing Guo[1] | Qizhi Zhou[8] | Jiaxin Yin[9] | Tao Hu[9] | Mei Lin[1] | Zhi Liang[10] | Hongmei Zhou[1]

[1]State Key Laboratory of Oral Diseases, Department of Oral Medicine, West China Hospital of Stomatology, Sichuan University, Chengdu, China

[2]Department of Oral Medicine, The Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China

[3]Department of Oral Medicine, The Affiliated Hospital of Qingdao University, Qingdao, China

[4]Department of Oral Medicine, School and Hospital of Stomatology, Wuhan University, Wuhan, China

[5]Department of Oral Pathology, West China Hospital of Stomatology, Sichuan University, Chengdu, China

[6]Department of Oral and Maxillofacial surgery, West China Hospital of Stomatology, Sichuan University, Chengdu, China

[7]Department of Oral Medicine, North Sichuan Medical College, Nanchong, China

[8]School of Acupuncture-Moxibustion and Tuina, Chengdu University of TCM, Chengdu, China

[9]Department of Preventive Dentistry, West China Hospital of Stomatology, Sichuan University, Chengdu, China

[10]Hefei National Laboratory for Physical Sciences at Microscale and School of Life Science, University of Science and Technology of China, Hefei, China

Correspondence
Hongmei Zhou and Mei Lin, State Key Laboratory of Oral Diseases, Department of Oral Medicine, West China Hospital of Stomatology, Sichuan University, Chengdu,

Abstract

Background: Despite their high accuracy to recognize oral potentially malignant disorders (OPMDs) with cancer risk, non-invasive oral assays are poor in discerning whether the risk is high or low. However, it is critical to identify the risk levels, since high-risk patients need active intervention, while low-risk ones simply need to be follow-up. This study aimed at developing a personalized computational model to predict cancer risk level of OPMDs and explore its potential web application in OPMDs screening.

Methods: Each enrolled patient was subjected to the following procedure: 1) personal information collection, 2) non-invasive oral examination, 3) oral tissue biopsy and 4) histopathological analysis, 5) treatment, and 6) follow-up. Patients were randomly divided into a training set (N = 159) and a test set (N = 107). Random forest was used to establish classification models. A baseline model (model-B) and a personalized model (model-P) were created. The former used the non-invasive scores only, while the latter was incremented with appropriate personal features.

Results: We compared the respective performance of cancer risk level prediction by model-B, model-P, and clinical experts. Our data suggested that all three have a similar level of specificity around 90%. In contrast, the sensitivity of model-P is beyond 80% and superior to the other two. The improvement of sensitivity by model-P reduced the misclassification of high-risk patients as low-risk ones. We deployed model-P in web.opmd-risk.com, which can be freely and conveniently accessed.

Conclusion: We have proposed a novel machine-learning model for precise and cost-effective OPMDs screening, which integrates clinical examinations, machine learning, and information technology.

Notes: In terms of modelling, the following steps can be detailed:
1) Personal information collection: labelled, ingest and aggregate data. Then the data shall be annotated.
2) Non-invasive oral examination: no instruments shall be introduced into the body, i.e.: vital staining, oral cytology, optical detection, saliva biomarkers detection, and image analysis
Train:Test is 60:40 so this is most likely a random classification, at least during the first layers

Wang, Yang and Wei are co-first authors and contributed equally to this work.

Trial registration – registry and number: The trial is registered in ClinicalTrials.gov (No. ChiCTR-DDT-13003221).

Sichuan 610041, China.
Emails: zhouhm@scu.edu.cn (H. Z.); linmei2k@163.com (M. L.)

Zhi Liang, Hefei National Laboratory for Physical Sciences at Microscale and School of Life Science, University of Science and Technology of China, Hefei, Anhui 230027, China.
Email: liangzhi@ustc.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 81772898, 81272962; Key Research and Development Program of Sichuan Province, Grant/Award Number: 2017SZ0104
Ministry of Science and Technology, Grant/Award Number: 2014CB910600;

KEYWORDS
cancer risk level prediction, non-invasive screening, oral potentially malignant disorders, personalized model, web application

1 | INTRODUCTION

Oral potentially malignant disorders (OPMDs) refer to all oral clinical presentations that carry a risk of malignant transformation, including oral leukoplakia (OLK), oral erythroplakia (OEK), oral lichen planus (OLP), and some refractory oral ulcerative/erosive lesions commonly found in clinical practice.[1,2] The estimated malignant transformation rate of OEK, OLK, and OLP is approximately 14.3%-50%, 0.13%-17.5%, and 0.4%-6.5%, respectively.[3-5] The majority of oral squamous cell carcinoma (OSCC) are originated from OPMDs.[4,6] Before developing to OSCC, most patients may experience with a long-term of OPMDs.[7] A specific lesion develops from hyperplasia to dysplasia at different degrees, then to carcinoma in situ, and finally to invasive carcinoma with an average course of 2.5-8.1 years.[2,8-10] Therefore, an effective supervision and intervention in patients with OPMDs are of utmost importance for cancer prevention.

Oral tissue biopsy and histopathological analysis are often considered as the gold standard for cancer risk assessment of OPMDs. However, since biopsy is an invasive assay, it may not be suitable for monitoring the chronic development of OPMDs when compared to non-invasive detection techniques.[11] Visually enhanced lesion (VEL) scope and toluidine blue (TB) staining are the most widely used non-invasive techniques for assessing cancer risk of OPMDs.[10] However, according to our long-term clinical practice, these two issues should be improved. Firstly, despite of their relatively high accuracy to determine lesions with cancer risk,[12] both the aforementioned non-invasive techniques are poor in discerning whether the cancer risk is high or low. Secondly, patients with different risk levels may present similar non-invasive results. It is of enormous importance to discriminate between high- and low-risk patients, since the high-risk ones need a "red alert" and active intervention, while low-risk patients simply need to be regularly observed.[13] In fact, individual information, such as age, gender, life style, and lesion conditions, plays important roles in the malignant transformation of OPMDs,[14] thus, integrating them may be a helpful complement for a successful result by non-invasive assays.

Therefore, in this study, these clinical challenges and obstacles were overcome for the first time by introducing a personalized machine-learning model to predict a patient's cancer risk level by integrating the two non-invasive oral assays with a patient's personal information. Our results showed that the model using the random forest (RF) technique could not only distinguish between high-risk and low-risk lesions with high sensitivity and specificity but it could also predict the risk in the future.

According to the American Cancer Society, cancer incidence is rapidly declining every year in the last 10 years, mainly because of the promotion of non-invasive screening.[15] Some studies showed that oral visual inspection at an interval of three years could reduce cancer-specific mortality of high-risk individuals by 26%~34%.[16,17] These results indicate the important role of screening in cancer prevention. Considering the widespread of the Internet through smartphones, we performed the personalized model in a web site called OPMDRisk (web.opmd-risk.com) for an easy accessibility and we explored its potential application and feasibility in OPMDs screening and monitoring.

2 | METHODS

2.1 | Patients

A total of 266 patients were enrolled from Department of Oral Medicine, West China Hospital of Stomatology, Sichuan University (called the SCU set hereafter), spanning from June 2013 to December 2016. Oral mucosal lesions of these patients were clinically suspected of being malignant, and therefore, a biopsy and consequent histopathological analysis were performed to confirm the cancer risk level. This study was conducted in accordance with ethical principles and approved by the Ethics Committee of our hospital (the trial was registered at ClinicalTrials.gov, number: ChiCTR-DDT-13003221). Written informed consent to participate in this study was obtained from all patients.

Chinese Clinical Trial Registry

Assess risk of opmd turning into oscc through improved non-invasive methods
Successful complement vs improved diagnosis

The inclusion criteria were as follows: (a) a lesion with a provisional clinical diagnosis of OLK, OEK, or OLP showing no response to conventional medical treatments for four weeks,[1,2] or (b) non-healing ulcer or erosive lesion suggesting a malignancy. Patients were excluded if they have a history of head and neck squamous cell carcinoma or malignant tumor in other body sites (see Appendix S1).

In addition to the SCU set, another set of patients was collected from Department of Oral Medicine, School and Hospital of Stomatology, Wuhan University, Wuhan, Hubei, China (called the WHU set). This set contains additional 38 patients serving as an independent test set.

2.2 | Study procedure (see Appendix S1)

Step 1. Personal information collection
Step 2. Non-invasive oral examination
Step 3. Biopsy and histopathological analysis
Step 4. Treatment and follow-up survey

The detailed information of each participant was recorded by using the specially designed registration table (Table S1). The semi-quantitative scoring systems of non-invasive oral examinations were presented in Figure 1. Briefly, we adopted semi-quantitative scoring systems by combining our long-term experience and the reported relationship between degree of fluorescence visualization loss or staining concentration and high-risk molecular profiles.[11,18] For VEL scope examination, the scoring system assigned 0 to green or white, 1 to gray, and 2 to dark. For TB staining, the scoring system assigned 0 to no staining, 1 to pale blue, and 2 to dark blue.

2.3 | Establishment of the machine-learning model

According to the collected personal details (eg, age and gender), conventional oral examination (eg, lesion site and clinical type) and non-invasive oral examination (ie, VEL scope and TB staining scores), each patient could be represented by multiple features and the corresponding histopathological grade of each patient represented the ground truth. Considering the clinical significance to discriminate between high- and low-grade patients, we adopted a binary classification of histopathological grades. The low grade includes non-dysplasia, mild dysplasia, and moderate dysplasia. And the high grade includes severe dysplasia, carcinoma in situ, and invasive carcinoma. The 266 patients of the SCU set were randomly divided into two groups, a training set and a test set. The former contained 3/5 of the data (159 patients), the latter contained the rest 2/5 (107 patients). The training set was used to train RF models using the randomForest package implemented in R environment with the default parameters.[19] The out-of-bag (OOB) error estimate of RF was used as a measure of model performance analysis in predicting cancer risk level of OPMDs. The test set was used as an additional independent data set to assess the prediction performance of trained models. We built two model types, a baseline model (model-B) and a personalized model (model-P). The former used the two non-invasive scores only, while the latter was incremented with additional personal features, such as age, gender, life habits, and lesion conditions. Two clinicians with more than 20 years' experience performed the same risk level prediction for all the 266 patients. The performance of predicting cancer risk level of OPMDs was compared to both models and experts by assessing sensitivity and specificity, where:

Sensitivity = No. of true positives/ (No. of true positives + No. of false negatives)
Specificity = No. of true negatives/ (No. of true negatives + No. of false positives)

The 38 patients of the WHU set were used as independent data to test the models.

2.4 | Long-term outcome prediction by the model

In order to investigate whether the personalized model could also predict the long-term risk of malignant development in a patient, we applied model-P to patients with a follow-up time of at least 3 years. Patients were then classified into a high-risk group and a low-risk group according to the corresponding prediction result. Then, the malignant transformation/ cancer relapse rate of these two groups was compared using Fisher's exact test.

Contingency calculation

2.5 | Web app development

not IaaS

The personalized prediction model was deployed on the Internet as a service for assessing the cancer risk level of a patient, which is freely accessible via the web app OPMDRisk (http://www.opmd-risk.com). The model was implemented with the random Forest package in R environment. The server was powered by PHP with MySQL support.

3 | RESULTS

3.1 | Patient information

A total of 304 patients were assessed for eligibility. Twenty-one of them did not meet the inclusion criteria, while 17 others who met the inclusion criteria refused to be subjected to biopsy. Finally, 266 patients including 116 men (43.61%) and 150 women (56.39%), with an age from 21 to 85 (average age 55.68 years) with disease duration between 3 weeks and 120 months (average duration 26.86 months) were enrolled (Figure 2). 0.75 of a month vs 10 years

Different characteristics of each patient were collected, including personal details, conventional oral examinations, and non-invasive oral examinations (Table 1). These inputs provide candidate

Categorical data vs Time Series: In this case many categorical data observed over time

Python default average trees: 10 VS R default at: 500 fDNN uses ReLU activation function

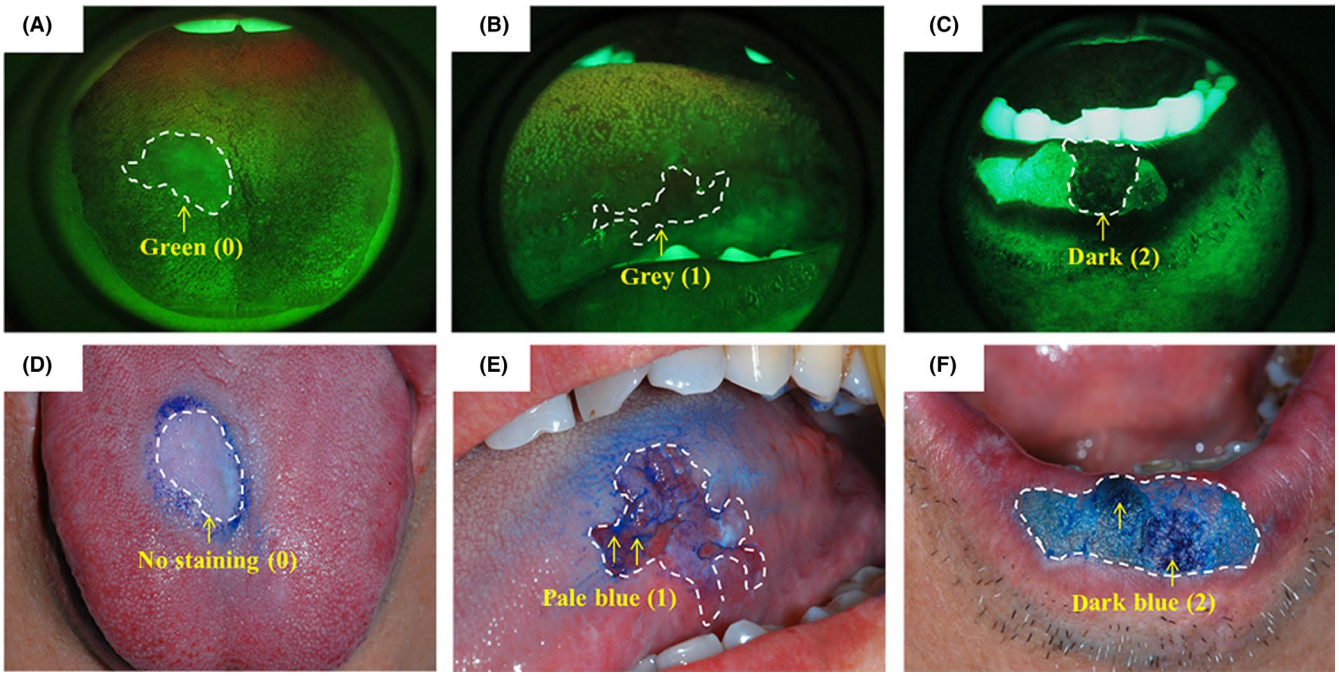

FIGURE 1 Semi-quantitative scoring system of non-invasive oral examinations. (A-C): VEL scope scoring. For visualization purposes, lesion regions were surrounded by dashed lines. Green (A), gray (B), and dark (C) scored as 0, 1, and 2, respectively. (D-F): TB staining scoring. The scoring system assigned 0 to no staining (D), 1 to pale blue (E), and 2 to dark blue (F). A false positive outside the lesion region was observed in (D), which resulted from the non-specific staining of the normal filiform papilla

predictors for subsequent modeling. The histopathological assessment of the oral tissue from each patient obtained through biopsy served as the ground truth label for model training and evaluation.

3.2 | Model building

RF is an ensemble learning method that has wide application in many fields.[19-21] In contrast to classical approaches for binary classification, such as logistic regression, RF was a relative new technique and has gained considerable popularity due to its impressive performance.[22,23] RF builds a large collection of decorrelated decision trees, each of which uses a different subset of all available factors to do prediction. Then, RF generates the final prediction by averaging the predicted results of all trees. In fact, a RF is just like a committee of experts, in which every expert (ie, a tree) gives its prediction based on its expertise (ie, the predictor variables used), and the committee makes the final prediction by combining the prediction of all the experts. Here, we used RF to model the cancer risk level of OPMDs given a patient's non-invasive oral examination and other personal information. The 266 patients from the SCU set were randomly divided into two sets, a training set with 3/5 of the data and a test set with the rest. Using the training set, two model types, model-B and model-P, were trained. Model-B, the baseline model, was based on the non-invasive oral examination only, namely the TB staining and VEL scope scores (Table 2). Besides the two non-invasive scores, model-P, the personalized model, included additional personal characteristics, such as lesion conditions, age, gender, and habits. In order

predictor accuracy vs model accuracy

to identify the subset of personal predictors that could improve the prediction accuracy of model-P over model-B for predicting cancer risk level of OPMDs, we began with an initial model-P containing all the predictors, and then we removed the less influential ones. For each predictor subset, we monitored the OOB error estimate of the model and investigated the predictor importance using the measure based on the Gini index. As regard predictors with similar importance measures, their combinations were explored. Finally, we obtained an optimal model-P that contains the two non-invasive scores and other four personal predictors, including the lesion's infiltration, its clinical type and site, and patient's age (Table 2).

Gini inequality calculation: how far is the predictor from drawing random predictions. python.numpy.random can be simulated using a Mersenne number: (2^n) -1

3.3 | Model evaluation

According to the OOB estimate based on the training data, the sensitivity and specificity of model-B for predicting cancer risk level of OPMDs were 80.00% and 92.31%, respectively (Table 3). Model-P improved the sensitivity to 85.45%, while keeping the same level of specificity (Table 3). The high specificity indicated that more than 90% of low-grade patients who were in a mild stage could be correctly predicted as low-risk by both models. In contrast, as shown by the sensitivity, the error of predicting high-grade patients as low-risk was higher. However, the error was reduced from 20.00% of model-B to 14.54% of model-P. Since the misclassification of high-grade patients might lead to delay in the treatment, the improvement of model-P by ~5.5 percentage points over model-B led to an improved outcome. The result indicated that the increment with appropriate personal information was helpful for improving model performance (Figure 3A-H).

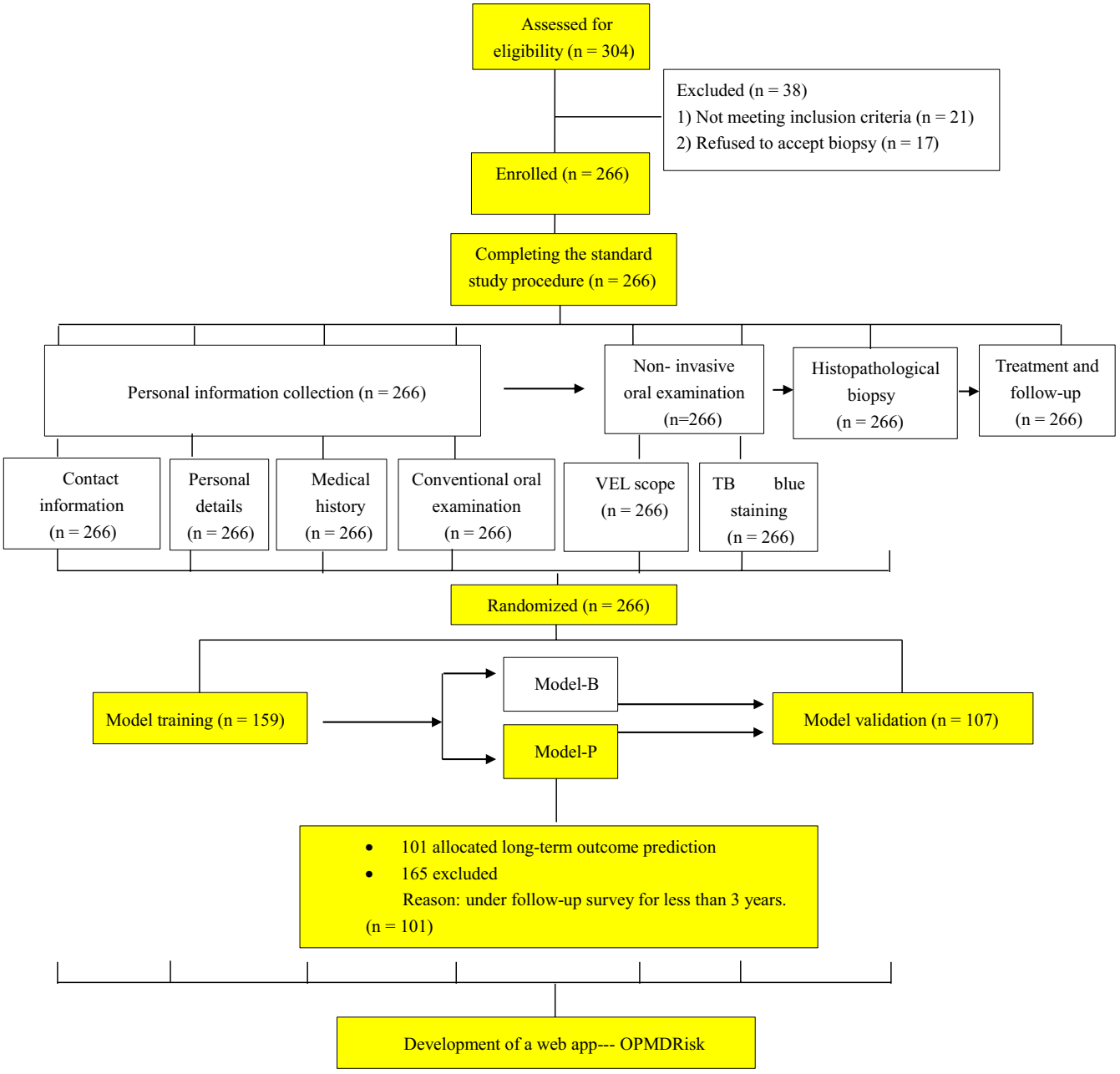

FIGURE 2 Flow diagram of the trial process

As an independent test, model-B and model-P were applied to the test set. The result consistent with the above analysis showed that model-P exhibited higher performance (Table 3). Both model-B and model-P achieved a high specificity of 91.78%, while the sensitivity of model-B was ~3 percentage points lower than the 82.35% of model-P.

Clinician experts were considered as an additional reference for model evaluation. We exposed the SCU set to the two clinician experts. The experts were asked to assess the cancer risk level of all patients based on the patients' personal details, and conventional and non-invasive oral examinations. By comparing the experts' predictions with the ground truth labels, the performance of the experts was determined. The experts achieved a sensitivity of 65.45% and a specificity of 92.31% for the training set, and a sensitivity of 79.41% and a specificity of 86.30% for the test set (Table 3), suggesting that model-B and model-P were superior to the experts evaluation thanks to the improved sensitivity and specificity of these two models in predicting cancer risk level of OPMDs.

Patients whose predictions were inconsistent among the experts, model-B and model-P were further investigated. Model-P effectively corrected many false negatives made by the experts and model-B (see Appendix S1).

To further test the generalization of the models, we also tested model-B and model-P using the WHU set. Similar to the above result, model-P showed a higher sensitivity than model-B (see Appendix S1).

TABLE 1 Characteristics of patients involved in this study

Personal details	
Age (y), mean ± SD	55.68 ± 11.99
Gender, n (%)	
Male	116 (43.61%)
Female	150 (56.39%)
Smoking history, n (%)	
Never	170 (63.91%)
Ever[a]	39 (14.66%)
Current[b]	57 (21.43%)
Drinking history, n (%)	
Never	168 (63.16%)
Ever[c]	61 (22.93%)
Current[d]	37 (13.91%)
Spicy food history, n (%)	
No	57 (21.43%)
Yes[e]	209 (78.57%)
Lesion conditions (conventional oral examinations)	
Site, n (%)	
Mouth commissure	3 (1.13%)
Lip	3 (1.13%)
Bucca	77 (28.95%)
Mouth floor	7 (2.63%)
Tongue venter	88 (33.08%)
Tongue dorsum	37 (13.91%)
Tongue margin	12 (4.51%)
Palate	14 (5.26%)
Gingiva	25 (9.40%)
Clinical type, n (%)	
Leukoplakia	197 (74.06%)
Erythroplakia	13 (4.89%)
Oral lichen planus	19 (7.14%)
Ulcer/ Erosive lesion	37 (13.91%)
Maximal width (mm), mean ± SD	12.29 ± 8.41
Maximal height (mm), mean ± SD	18.66 ± 11.62
Keratinization (score), n (%)	
0	41 (15.41%)
1	91 (34.21%)
2	85 (31.96%)
3	49 (18.42%)
Infiltration, n (%)	
No	223 (83.83%)
Yes	43 (16.17%)
Texture, n (%)	
Soft	167 (62.78%)
Moderate hard	84 (31.58%)
Hard	15 (5.64%)
Non-invasive oral examinations	
VEL scope (score)[f], n (%)	
0	89 (33.46%)
1	111 (41.73%)
2	66 (24.81%)
TB staining (score)[f], n (%)	
0	124 (46.62%)
1	77 (28.95%)
2	65 (24.43%)
Biopsy	
Histological grade[g], n (%)	
High-grade lesions[h]	89 (33.46%)
Low-grade lesions[i]	177 (66.54%)

[a]Ever smokers: those who had a lifetime history of smoking/ chewing more than 100 cigarettes, cigars, or pipes, but stopped for at least 1 year prior to the trial.

[b]Current smokers: those who had a lifetime history of smoking/ chewing more than 100 cigarettes, cigars, or pipes, and were currently smoking/ chewing "every day" or "some days" prior to the trial.

[c]Ever drinkers: those who drank twice or more beer, white wine, red wine, or liquor/ spirits per week for one year, but stopped for at least 1 year prior to the trial.

[d]Current drinker: those who drank twice or more beer, white wine, red wine, or liquor/ spirits per week for one year, and were currently drinking "every day" or "some days" prior to the trial.

[e]"Yes": patients who ate spicy food 10 times or more per week at least for 6 months prior to the trial.

[f]See Figure 1.

[g]On the basis of the histological classification criteria recommended by the World Health Organization.

[h]High-grade lesions: lesions with a pathological manifestation of severe dysplasia/carcinoma in situ/invasive carcinoma.

[i]Low-grade lesions: lesions with a pathological manifestation of no/ mild/ moderate dysplasia.

3.4 | Predictor importance for risk prediction

To learn the relative contribution of each patient's input variable in predicting the cancer risk of OPMDs, predictor importance was investigated using the mean decrease in Gini index for each variable, relative to the largest. The first and second most influential variables were VEL scope and TB staining scores (15.57 and 13.92, respectively). The observation was consistent with our clinical experience and the performance of model-B. Lesion conditions, including infiltration, clinical type, site, texture, keratinization, and minimal and maximal dimensions (6.74, 3.48, 3.49, 3.61, 3.94, 4.77, and 4.97, respectively), had a considerable impact. Degree of infiltration was the dominant lesion-related variable. Although they were influential, using all the lesion-related variables did not improve the model performance. In fact, only a proper predictor subset would lead to better prediction accuracy for cancer risk level of OPMDs. Age was another important

TABLE 2 Factors in Model-B and Model-P

	Non-invasive oral examination		Conventional oral examination (lesion conditions)			Personal details
	VEL scope	TB staining	Clinical type	Site	Infiltration	Age
Model-B	√	√	–	–	–	–
Model-P	√	√	√	√	√	√

TABLE 3 Cancer risk level prediction by model-B, model-P and experts based on the SCU set

	Prediction		
Based on the training set[a]	Low risk	High risk	Performance
Model-B			
Histological grade			
Low grade	96 (true negative)	8 (false positive)	Specificity 92.31%
High grade	11 (false negative)	44 (true positive)	Sensitivity 80.00%
Model-P			
Histological grade			
Low grade	96 (true negative)	8 (false positive)	Specificity 92.31%
High grade	8 (false negative)	47 (true positive)	Sensitivity 85.45%
Experts			
Histological grade			
Low grade	96 (true negative)	8 (false positive)	Specificity 92.31%
High grade	19 (false negative)	36 (true positive)	Sensitivity 65.45%
	Prediction		
Based on the test set	**Low risk**	**High risk**	**Performance**
Model-B			
Histological grade			
Low grade	67 (true negative)	6 (false positive)	Specificity 91.78%
High grade	7 (false negative)	27 (true positive)	Sensitivity 79.41%
Model-P			
Histological grade			
Low grade	67 (true negative)	6 (false positive)	Specificity 91.78%
High grade	6 (false negative)	28 (true positive)	Sensitivity 82.35%
Experts			
Histological grade			
Low grade	63 (true negative)	10 (false positive)	Specificity 86.30%
High grade	7 (false negative)	27 (true positive)	Sensitivity 79.41%

[a]The performance of model-B and model-P was obtained using OOB estimate of RF models.

predictor that could improve model performance, with score of 5.62. Gender and personal habits, such as drinking, smoking, and spicy food consumption, showed no clear influence (0.77, 1.49, 1.09, and 0.99, respectively).

3.5 | Long-term outcome prediction

A total of 101 patients in the SCU set were under follow-up survey for at least 3 years. These patients provided us an opportunity to investigate the ability of the personalized model for predicting long-term outcome. Model-P was applied to these patients, and 30 and 71 of them were predicted as high-risk and low-risk, respectively. Although both groups had similar time course to develop cancer or experience cancer relapse (42.9 ± 8.66 vs 43.36 ± 7.10, P = .8), the difference in the rate of malignant transformation of OPMDs/ cancer relapse was significant (P = .018). For the high-risk group, 26.67% (8 out of 30) of patients developed cancer. In contrast, the low-risk group had a lower cancer development rate (7.04% or 5 out of 71). Therefore, the significant difference indicated that the personalized model could also serve as an indicator for long-term risk assessment (Figure 3I-P).

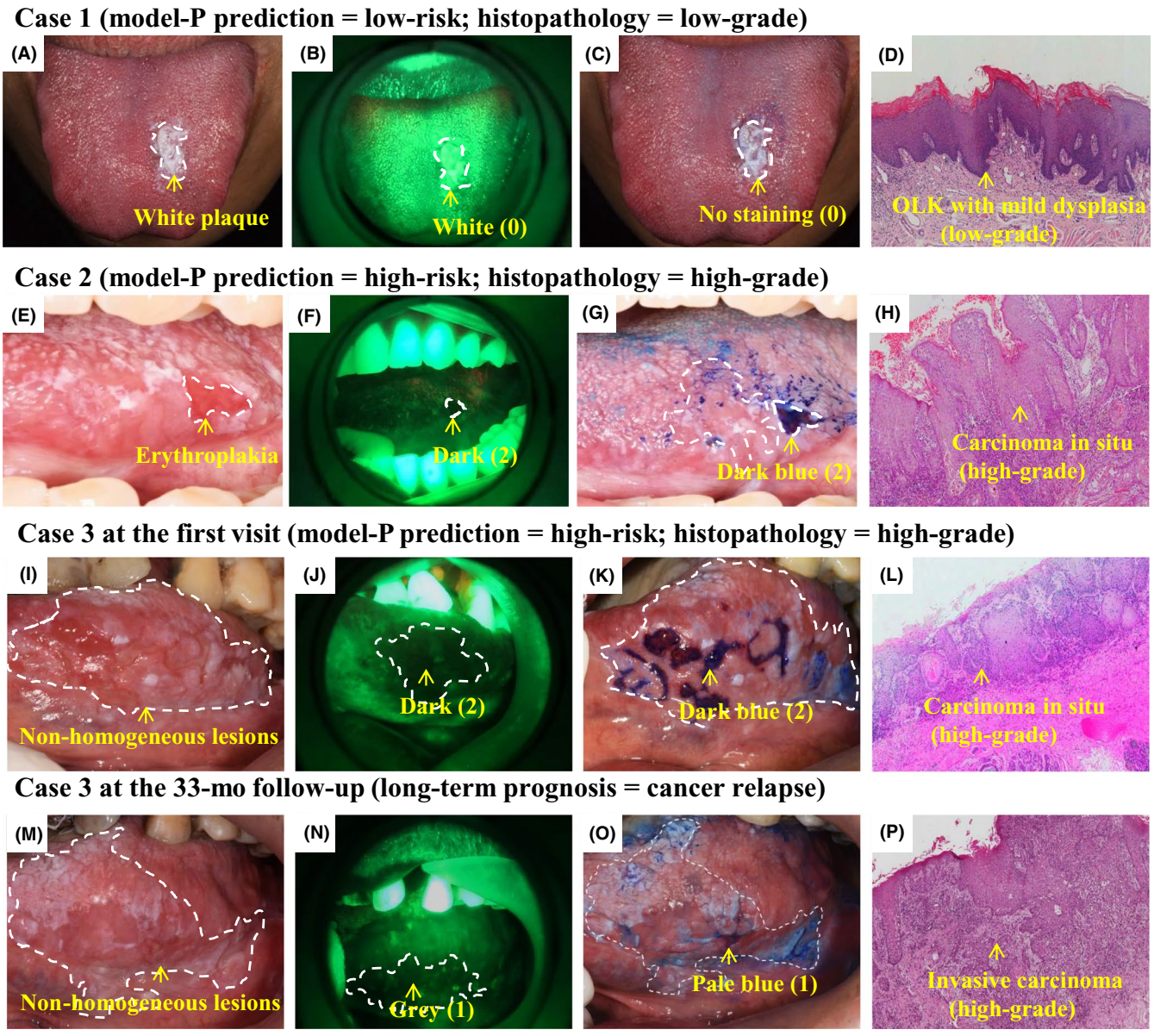

FIGURE 3 Present and long-term cancer risk prediction of model-P for OPMDs. (A-D): Case 1. A localized white plaque present on the tongue dorsum (A). VEL scope (B) and TB staining (C) examinations showing a white fluorescence (score = 0) and no staining (scored = 0), respectively. Model-P indicated a low-risk for malignant transformation, consistent with the examination of histopathological analysis of the biopsy identifying an OLK with mild dysplasia (low-grade) (40× magnification) (D). (E-H): Case 2. An irregular non-homogeneous lesions with a subtriangular erythroplakia observed in left tongue venter (E). Both VEL scope (F) and TB staining (J) scores were 2. Model-P indicated a high-risk for malignant transformation. The histopathology confirmed it as a high-grade lesion (carcinoma in situ) (40× magnification) (H). (I-L): Case 3 at first visit. Widespread non-homogeneous lesions, including the scattered white plaque, multiple atrophy, and erosion, were visible in the left tongue venter (I). Both VEL scope (J) and TB staining (K) scores were 2. Model-P indicated a high-risk for malignant transformation. The histopathological analysis of the biopsy confirmed the high-grade lesions as a carcinoma in situ (40× magnification) (L). Therefore, the patient was immediately subjected to an extended resection. (M-P): Case 3 at the follow-up of the 33rd month. New extensive non-homogeneous lesions were observed, including the scattered white plaque and erosion in the left tongue venter where the former lesions were completely removed (M). Both VEL scope (N) and TB staining (O) scored 1. The second pathological analysis of the biopsy revealed the high-grade lesion as invasive carcinoma (40× magnification) (P)

3.6 | Risk levels prediction service via the OPMDRisk web app

Considering the important role of non-invasive screening for cancer prevention[15] and the widespread of Internet through smart phones, we developed a web app called OPMDRisk, so that the risk level prediction for OPMDs could be freely and conveniently accessed all around the world. Any user can predict the cancer risk level of OPMDs by providing patient's information such as TB staining and VEL scope scores, and lesion conditions obtained from conventional oral examination, and personal details.

4 | DISCUSSION

Currently, VEL scope and TB staining are the most commonly used non-invasive screening techniques. However, two issues were found during our clinical practice. On one hand, significantly different interpretation might exist for the same non-invasive result by different clinicians because of the absence of consistent evaluation criteria. Hence, we developed two scoring systems for VEL scope and TB staining techniques (Figure 1). The two scoring systems defined concrete criteria to evaluate lesions, aiming at reducing the divergence due to subjective assessment.

On the other hand, although both non-invasive detection techniques could differentiate lesions with high accuracy,[24-26] they are poor in predicting the high or low level of cancer risk.[27] Many studies showed that individual factors could be related to the cancer development of OPMDs,[9,28] and thus they might be helpful for improving prediction accuracy of cancer risk level. In light of successful applications of individualized prediction in other diseases, such as cardiovascular disease and prostate cancer,[14] we used for the first time machine-learning techniques to develop a personalized model by integrating various individual variables and the two non-invasive predictors.

The cases with inconsistent predictions from the experts, model-B and model-P were analyzed. In fact, most of the misclassified patients exhibited weakly positive signals for both VEL scope and TB staining scores. In other words, these patients were on the "boundary" of the two classes, high-risk and low-risk, being difficult for the prediction of the risk level by the experts or model-B using the two non-invasive scores only. As shown above, incorporating suitable individual factors became helpful in resolving the uncertainty in this case. The preliminary analysis of the generalization of the models showed that the models might be applicable to other populations to some extent, at least for populations of central China (see Appendix S1). In addition to the two non-invasive scores, the other four personal predictors, including the lesion's infiltration, its clinical type and site, and patient's age, are wide spectrum factors which are non-specific across populations and may be a major reason for the applicability of the models. In addition, the personalized model not only showed the ability of predicting the risk level of a patient in the current time, it can also indicate the risk in the future. However, the main limitation of this study is that currently the model was based on a limited amount of data; we expect feedbacks from users, which would be very helpful to improve the model.

It has been demonstrated that promotion of cancer screening in the general population contributes to early detection of cancer risk and initiation of adequate intervention for patients with OPMDs and oral cancer.[16,17,29] Besides, the US 2017 and 2018 cancer reports also pointed out that promoting effective screening techniques could significantly reduce cancer incidence and mortality.[15,30] Inspired by these evidences, we believe that it would be of benefit to promote the application of the personalized computational model established in this study. It might be of important practical significance to help reducing the incidence of OSCC and improve patients' survival rate.

In order to make the personalized prediction model accessible to clinicians all over the world, we developed the OPMDRisk web app. OPMDRisk provides a simple and user-friendly interface to the personalized prediction model. Following standardized conventional and non-invasive oral examinations, the user submits the scores and individual information into the app. The app will reveal the predicted cancer risk level. We hope this app can serve as a convenient and cost-effective tool for cancer risk level prediction for OPMDs. Furthermore, with feedbacks and cooperation of the community, it could evolve into a public data hub for OPMDs and enables the development of more advanced tools in the future.

In conclusion, a novel model for precise and cost-effective OPMDs/OSCC screening was proposed, which integrated clinical examinations, machine-learning, and information technology. We hope that our work could provide the basis to establish standard procedures for OPMDs/OSCC screening and monitoring.

ACKNOWLEDGEMENTS

This work was supported by National Natural Science Foundation of China No. 81772898, 81272962 and Key Research and Development Program of Sichuan Province No. 2017SZ0104 to Hongmei Zhou, Ministry of Science and Technology 2014CB910600 to Zhi Liang.

CONFLICT OF INTEREST

The authors do not have any conflict of interest to declare.

ORCID

Hongmei Zhou https://orcid.org/0000-0003-2985-3536

REFERENCES

1. Warnakulasuriya S, Johonson NW, van der Waal I. Nomenclature and classification of potentially malignant disorders of the oral mucosa. *J Oral Pathol Med*. 2007;36:575-580.
2. Van der Waal I. Potentially malignant disorders of the oral and oropharyngeal mucosa; terminology, classification and present concepts of management. *Oral Oncol*. 2009;45:317-323.
3. Reichart PA, Philipsen HP. Oral erythroplakia. A review. *Oral Oncol*. 2005;41:551-561.
4. Qin GZ, Park JY, Chen SY, Lazarus P. A high prevalence of p53 mutations in pre-malignant oral erythroplakia. *Int J Cancer*. 1999;80:345-348.
5. Holmstrup P, Vedtofte P, Reibel J, Stoltze K. Long-term treatment outcome of oral premalignant lesions. *Oral Oncol*. 2006;42:461-474.
6. Ho PS, Chen PL, Warnakulasuriya S, Shieh TY, Chen YK, Huang IY. Malignant transformation of oral potentially malignant disorders in males: a retrospective cohort study. *BMC Cancer*. 2009;30(9):260.
7. Thomson PJ, Goodson ML, Smith DR. Profiling cancer risk in oral potentially malignant disorders-A patient cohort study. *J Oral Pathol Med*. 2017;46:888-895.
8. Wei CL, Gao QH, Wu LY, et al. Non/micro-invasive clinicopathologic methods in the assessment of oral leukoplakia multistep carcinogenesis: a case report. *IntJ Clin Exp Patho*. 2016;9:9687-9693.
9. Zhang L, Williams M, Poh CF, et al. Toluidine blue staining identifies high-risk primary oral premalignant lesions with poor outcome. *Cancer Res*. 2005;65:8017-8021.
10. Silverman S Jr, Gorsky M, Lozada F. Oral leukoplakia and malignant transformation. A follow-up study of 257 patients. *Cancer*. 1984;53:563-568.

11. Maeda H, Kameyama Y. Effect of excisional wounding on DMBA-induced hamster tongue carcinogenesis. *J Oral Pathol*. 1986;15:21-27.
12. Pallagatti S, Sheikh S, Aggarwal A, et al. Toluidine blue staining as an adjunctive tool for early diagnosis of dysplastic changes in the oral mucosa. *J Clin Exp Dent*. 2013;5:e187-e191.
13. Kujan O, Oliver RJ, Khattab A, Roberts SA, Thakker N, Sloan P. Evaluation of a new binary system of grading oral epithelial dysplasia for prediction of malignant transformation. *Oral Oncol*. 2006;42:987-993.
14. Levy WC, Mozaffarian D, Linker DT, et al. The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation*. 2006;113:1424-1433.
15. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA Cancer J Clin*. 2017;67:7-30.
16. Sankaranarayanan R, Ramadas K, Thomas G, et al. Effect of screening on oral cancer mortality in Kerala, India: a cluster-randomised controlled trial. *Lancet*. 2005;365:1927-1933.
17. Chuang SL, Su WW, Chen SL, et al. Population-based screening program for reducing oral cancer mortality in 2,334,299 Taiwanese cigarette smokers and/or betel quid chewers. *Cancer*. 2017;123:1597-1609.
18. Goodson M, Sloan P, Thomson P. Interobserver agreement for velscopegrading of normal and pathological oral mucosa. *Oral Oncol*. 2013;49:S98.
19. Breiman L. Random forests. *Mach Learn*. 2001;45:5-32.
20. Criminisi A, Shotton J. Decision forests for computer vision and medical image analysis. In: Kang SB, ed. *Advances in Computer Vision and Pattern Recognition*. London: Springer; 2013.
21. Zhang H, Yu CY, Singer B. Cell and tumor classification using gene expression data: construction of forests. *Proc Natl Acad Sci USA*. 2003;100:4168-4172.
22. Couronne R, Probst P, Boulesteix A. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics*. 2018;19:270.
23. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. London, UK: Springer; 2009.
24. Petruzzi M, Lucchese A, Nardi GM, et al. Evaluation of autofluorescence and toluidine blue in the differentiation of oral dysplastic and neoplastic lesions from non dysplastic and neoplastic lesions: a cross-sectional study. *J Biomed Opt*. 2014;19:76003.
25. Scheer M, Neugebauer J, Derman A, Fuss J, Drebber U, Zoeller JE. Autofluorescence imaging of potentially malignant mucosa lesions. *Oral SurgOral Med Oral Pathol Oral Radiol Endod*. 2011;111:568-577.
26. Farah CS, McIntosh L, Georgiou A, McCullough MJ. Effcacy of tissue autofluorescence imaging (VELScope) in the visualization of oral mucosal lesions. *Head Neck*. 2012;34:856-862.
27. Rashid A, Warnakulasuriya S. The use of light-based (optical) detection systems as adjuncts in the detection of oral cancer and oral potentially malignant disorders: a systematic review. *J Oral Pathol Med*. 2015;44:307-328.
28. Goldenberg D, Brooksby C, Hollenbeak CS. Age as a determinant of outcomes for patients with oral cancer. *Oral Oncol*. 2009;45:e57-e61.
29. Amarasinghe HK, Johnson NW, Lalloo R, Kumaraarachchi M, Warnakulasuriya S. Derivation and validation of a risk-factor model for detection of oral potentially malignant disorders in populations with high prevalence. *Br J Cancer*. 2010;103:303-309.
30. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018.*CA Cancer J Clin*. 2018;68:7-30.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Wang X, Yang J, Wei C, et al. A personalized computational model predicts cancer risk level of oral potentially malignant disorders and its web application for promotion of non-invasive screening. *J Oral Pathol Med*. 2020;49:417–426. https://doi.org/10.1111/jop.12983