

RESEARCH

Open Access



Application of machine learning for the analysis of peripheral blood biomarkers in oral mucosal diseases: a cross-sectional study

Huiyu Yao¹, Zixin Cao¹, Liangfu Huang¹, Haojie Pan¹, Xiaomin Xu¹, Fucai Sun¹, Xi Ding^{1*} and Wan Wu^{1*}

Abstract

Background Oral mucosal lesions are widespread globally, have a high prevalence in clinical practice, and significantly impact patients' quality of life. However, their pathogenesis remains unclear. Recent evidences suggested that hematological parameters may play a role in their development. Our study investigated the differences in humoral immune indexes, serum vitamin B levels, and micronutrients among patients with oral mucosal lesions and healthy controls. Additionally, it evaluated a Random Forest machine learning model for classifying various oral mucosal diseases based on peripheral blood biomarkers.

Methods We recruited 237 patients with recurrent aphthous ulcers (RAU), 35 with oral lichen planus (OLP), 67 with atrophic glossitis (AG), 35 with burning mouth syndrome (BMS), and 82 healthy controls. Clinical data were analyzed by SPSS 24 software. Serum levels of immunoglobulins (IgG, IgA, IgM), complements (C3, C4), vitamin B (VB1, VB2, VB3, VB5), serum zinc (Serum Zn), serum iron (Serum Fe), unsaturated iron-binding capacity (UIBC), total iron-binding capacity (TIBC), and iron saturation (Iron Sat) were measured and compared among groups. A Random Forest model was applied to analyze a dataset comprising 319 samples with eight key biomarkers.

Results Significant differences were observed between the oral mucosal diseases groups and controls in the serum levels of VB2, VB3, VB5, zinc, iron, TIBC, and Iron Sat. Specifically, serum levels of VB2 and VB3 were significantly higher in patients compared to controls ($*p < 0.05$), while levels of VB5, Serum Zn, Serum Fe, TIBC, and Iron Sat were significantly lower ($*p < 0.05$). No significant differences were found for C3, C4, IgG, IgM, IgA, VB1, and UIBC. The optimized Random Forest model demonstrated high performance, and effectively classified different disease groups, though some overlap between groups was noted. Feature importance analysis, based on the Mean Decrease Accuracy and Gini Index, identified VB2, VB3, Serum Fe, TIBC, and Serum Zn as key biomarkers, indicating their potential in distinguishing oral mucosal diseases.

Conclusion Our study identified significant associations between the contents of VB2, VB3, VB5, Serum Fe, Serum Zn, and other micronutrients and oral mucosal lesions. It suggested that regulating these micronutrient levels could be essential for preventing and curing such lesions. The Random Forest model demonstrated high accuracy (94.68%)

*Correspondence:

Xi Ding
dingxi@wzhospital.cn
Wan Wu
13857788811@163.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

in classifying disease groups, emphasizing the potential of machine learning to enhance diagnostic precision in oral mucosal diseases. Future research should focus on validating these findings in larger cohorts and exploring alternative machine-learning algorithms to improve diagnostic accuracy further.

Keywords Oral mucosal lesions, Humoral immunity, Vitamin B, Micronutrients, Machine learning, Random forest model

Introduction

Oral mucosal lesions encompass a wide spectrum of conditions affecting the soft tissues and mucosa of the oral cavity, which could cause a few sufferings, especially ulceration, and erosion. Recurrent aphthous ulcers (RAU), oral lichen planus (OLP), atrophic glossitis (AG), and burning mouth syndrome (BMS) were some of the highest incidence rates of oral mucosal lesions in the latest epidemiological survey in China, which tortured more than 80 million people [1]. It is common for these lesions to alternate, overlap, or coexist at different stages of their progression, causing clinical complexity [2]. RAU is one of the most common ulcerative lesions, with a global prevalence rate of approximately 20%, which could generate discomfort and reduce the life quality of patients [3, 4]. OLP was estimated to affect up to 2% of the general population and was classified as an oral potentially malignant disorder in 2017, with the potential to progress to oral squamous cell carcinoma (OSCC) [5]. AG impacted about 240 million people globally, with a higher prevalence in middle-aged and elderly individuals [6, 7]. BMS affects approximately 90 to 120 individuals per 100,000, while females were seven times more likely to be diagnosed compared with males with an average onset age of around 59 [8]. Oral mucosal diseases above would cause discomfort and reduce their appetite, resulting in nutritional deficiencies and immune system disorders [9, 10]. Meanwhile, several literatures indicated that deficiencies in micronutrients and vitamin B were observed in patients with these oral mucosal diseases [11], which could also be frequently encountered in clinical practice.

The etiology and pathogenesis of oral mucosal lesions remained unclear and multifaceted, marked by significant individual variability. While various pathogenic factors have been implicated, including immune dysregulation, genetic predisposition, systemic conditions and environmental influences, a unified understanding of their roles has yet to be established [12]. Among these factors, hematological parameters have been considered possible etiological contributors. The blood test is a routine and accessible diagnostic tool commonly used in outpatient service, offering a convenient and reliable method for evaluating nutritional status. Immunological mechanisms were considered to play a critical role in the development of these lesions [13]. Several studies have demonstrated the correlation between oral mucosal lesions and alterations in humoral immune markers.

For example, elevated levels of IgM, IgG, IgA, and IgE have been observed in patients with RAU and other oral mucosal conditions, suggesting potential disruptions in humoral immune function. Furthermore, a significant inverse correlation has been reported between Th17 cell activity and IgA expression, indicating abnormal humoral immunity potentially related to T cell dynamic. Additionally, abnormalities in hematological indices, such as mean erythrocyte volume, mean hemoglobin, serum iron (Serum Fe), vitamin B12, and folic acid levels, have been detected in patients with RAU and AG compared to healthy controls [14]. Moreover, deviations in trace elements, folic acid, and vitamin B12 levels are closely associated with a range of oral mucosal lesions. It is becoming a rising problem whether laboratory examinations could improve diagnostic accuracy or provide insights into the underlying etiological mechanisms of these conditions.

In this study, we sought to evaluate the differences in peripheral blood levels of humoral immunity markers, micronutrients, and B vitamins between patients with oral mucosal lesions and healthy controls. Furthermore, we assessed the importance of differential biomarkers in classifying diseases by the Random Forest model. In clinical research, the Random Forest model, with its built-in feature importance evaluation [15], allowed the model trained on clinical data to identify biomarkers that exhibit more significant changes in the disease. Additionally, the model could recognize combinations of these biomarkers through decision trees, helping to distinguish between different categories, which showcased its potential in disease prediction [16].

Materials and methods

Study design and population

This study included 319 participants, including 237 patients with oral mucosal lesions and 82 healthy controls. Among these patients, 100 had recurrent aphthous ulcers (RAU), 35 had oral lichen planus (OLP), 67 had atrophic glossitis (AG), and 35 had burning mouth syndrome (BMS). The participants were recruited from the Department of Stomatology at The First Affiliated Hospital of Wenzhou Medical University between November 2021 and March 2023. Healthy individuals ($n = 82$) were recruited from the hospital's physical examination center as the control group. The flowchart of the study was displayed in Fig. 1.

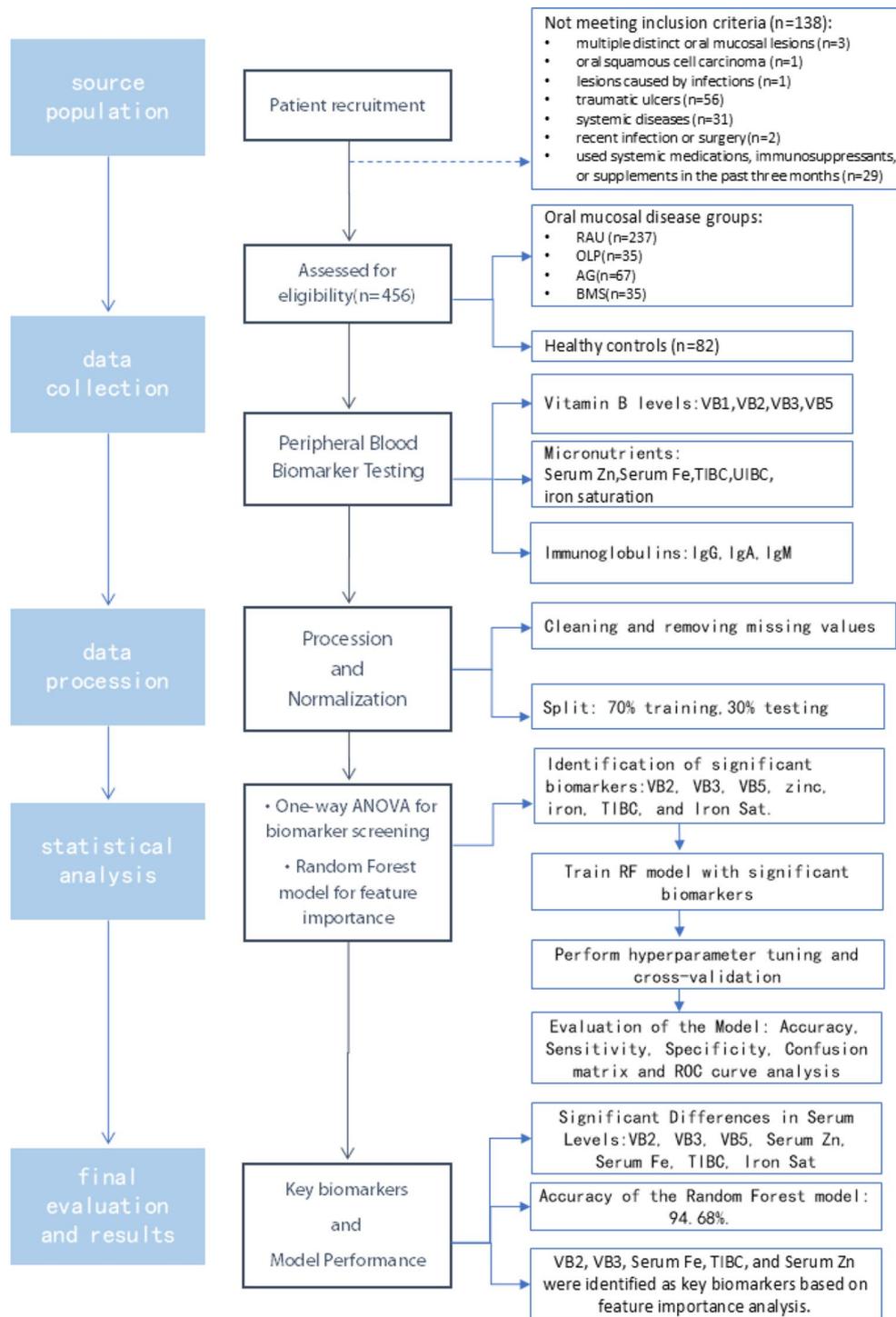


Fig. 1 STROBE flowchart of the study

Patients with oral mucosal lesions were identified through clinical presentation and medical history. The diagnosis was confirmed by the same physician based on clinical findings and supplementary examinations, including laboratory tests and pathological examination. Healthy controls were selected based on the absence of

symptoms and systemic diseases, and all participants provided written informed consent.

The study was approved by the Ethics Committee in Clinical Research (ECCR) of the First Affiliated Hospital of Wenzhou Medical University (No. KY2024-R233). All procedures were performed in accordance with the Declaration of Helsinki.

Inclusion and exclusion criteria

The study included adults diagnosed with specific oral mucosal diseases who agreed to provide blood samples, while the control group consisted of healthy adults without any reported symptoms or systemic diseases. Patients were included aged 18–80 years and diagnosed with one of the following oral mucosal conditions: RAU, OLP, AG, or BMS. The diagnosis criteria are based on the International Classification of Diseases 11th Revision (ICD-11) and the ESMO Clinical Practice Guidelines [17, 18]. All participants provided blood samples early in the morning after an overnight fast. All blood samples were collected by professionals at the blood collection center of the outpatient department at The First Affiliated Hospital of Wenzhou Medical University.

The exclusion criteria for participants in the study group were as follows:

1. Individuals presenting with multiple distinct oral mucosal lesions simultaneously.
2. Individuals with oral mucosal lesions of an ulcerative nature, such as oral squamous cell carcinoma.
3. Individuals with lesions caused by infectious agents, such as bacteria or fungi, or with conditions such as AIDS-related oral mucosal lesions.
4. Individuals with specific causes of mucosal lesions, such as traumatic ulcers.
5. Individuals with systemic diseases.
6. Individuals with a history of acute infection or recent surgical procedures.
7. Individuals who had used systemic medications, immunosuppressants, micronutrient supplements, or nutritional supplements within the last three months.

Medical and laboratory data measurements

All biochemical parameters were analyzed at the Medical Laboratory Center of The First Affiliated Hospital of Wenzhou Medical University, following standardized quality control protocols. The parameters included vitamin B1 (VB1), vitamin B2 (VB2), vitamin B3 (VB3), and other relevant biomarkers. The high-performance liquid chromatography (HPLC) and ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) systems (API3200MDTM, Shimadzu-LC20-AD, AB Sciex-API-3) were utilized to quantify vitamin B1 (VB1), vitamin B2 (VB2), vitamin B3 (VB3), and vitamin B5 (VB5). A 200MD kit was employed for analyzing water-soluble vitamins. Immunoglobulins (IgG, IgA, IgM), complement C3, and complement C4 were measured using immunoturbidimetry (LX20PRO automatic biochemical analyzer, Beckman, USA; reagents provided by Zhejiang Ilikang Biotechnology Co.). Serum zinc (Serum Zn), serum iron (Serum Fe), unsaturated

iron-binding capacity (UIBC), total iron-binding capacity (TIBC), and iron saturation (Iron Sat) were determined using the azo-arsenic III method.

Statistical analyses

Data were analyzed using the SPSS software (version 24.0, IBM Corporation). All data were initially assessed for normality using the Shapiro-Wilk test. For normally distributed data, means and standard deviations (mean \pm SD) were calculated. One-way ANOVA was used for comparing means among multiple groups. If significant differences were found, post-hoc pairwise comparisons were performed using the Least Significant Difference (LSD) test. Non-normally distributed data were reported as medians and interquartile ranges [m (Q1, Q3)], and the Kruskal-Wallis H test was used for multiple group comparisons, with the Bonferroni correction applied for pairwise comparisons. Categorical data were analyzed using the chi-square (χ^2) test. A p-value below 0.05 was regarded as indicative of statistical significance for all tests.

Machine learning

Variables and relationships

The Group variable (RAU, OLP, AG, BMS, and healthy controls) was designated as the dependent variable, representing disease classification. Independent variables included serum biomarkers (VB2, VB3, VB5, Serum Zn, Serum Fe, TIBC, Iron Sat) that showed significant differences between groups ($*p < 0.05$) in statistical analyses. These biomarkers were hypothesized to reflect underlying pathophysiological mechanisms, such as micronutrient deficiency and immune dysregulation, which may influence disease susceptibility and presentation.

Model selection rationale and data preprocess

The Random Forest (RF) model was selected based on its demonstrated superiority in handling high-dimensional, non-linear biological data and its embedded feature selection capability. Data preprocessing involved removing missing values and converting categorical variables into appropriate formats for machine learning analysis. The dataset was divided into a training set (70%) and a testing set (30%) using stratified sampling to maintain class distribution.

Feature selection and hyperparameter configuration

The model's hyperparameters, such as the number of trees and the maximum depth, were optimized through 5-fold cross-validation to enhance model performance. The Random Forest was implemented with $n_{tree} = 500$ to ensure model stability. The m_{try} parameter was optimized via cross-validation, set to 2 after evaluating the square root of the number of features. Tree depth was not

Table 1 Demographic distribution of the patients in different oral mucosal lesions (n, %)

General information		control(n=82)	RAU(n=100)	OLP(n=35)	AG(n=67)	BMS(n=35)	H/ χ^2 value	P value
Gender	male	36(43.90%)	46(46.00%)	15(42.86%)	15(22.39%)	10(28.57%)	12.453	0.014*
	female	46(56.10%)	54(54.00%)	20(57.14%)	52(77.61%)	25(71.43%)		
Age(Year)		59.00(50.00, 68.00)	44.00(18.00, 70.00)	53.00(36.00, 70.00)	55.00(36.00, 74.00)	58.00(44.00, 72.00)	33.173	<0.001*
Season	spring	25(30.49%)	22(22.00%)	14(40.00%)	25(37.31%)	8(22.86%)	16.942	0.002*
	summer	42(51.21%)	26(26.00%)	5(14.28%)	28(41.79%)	11(31.42%)		
	autumn	6(7.32%)	28(28.00%)	8(22.86%)	4(5.97%)	12(34.29%)		
	winter	9(10.98%)	24(24.00%)	8(22.86%)	10(14.93%)	4(11.42%)		

Note: *Age does not follow a normal distribution and is represented by the 50th percentile (25th percentile, 75th percentile). Kruskal-Wallis H test was employed to perform multiple group comparisons, while a Bonferroni correction test was used to perform pairwise comparisons between groups. * $p < 0.05$

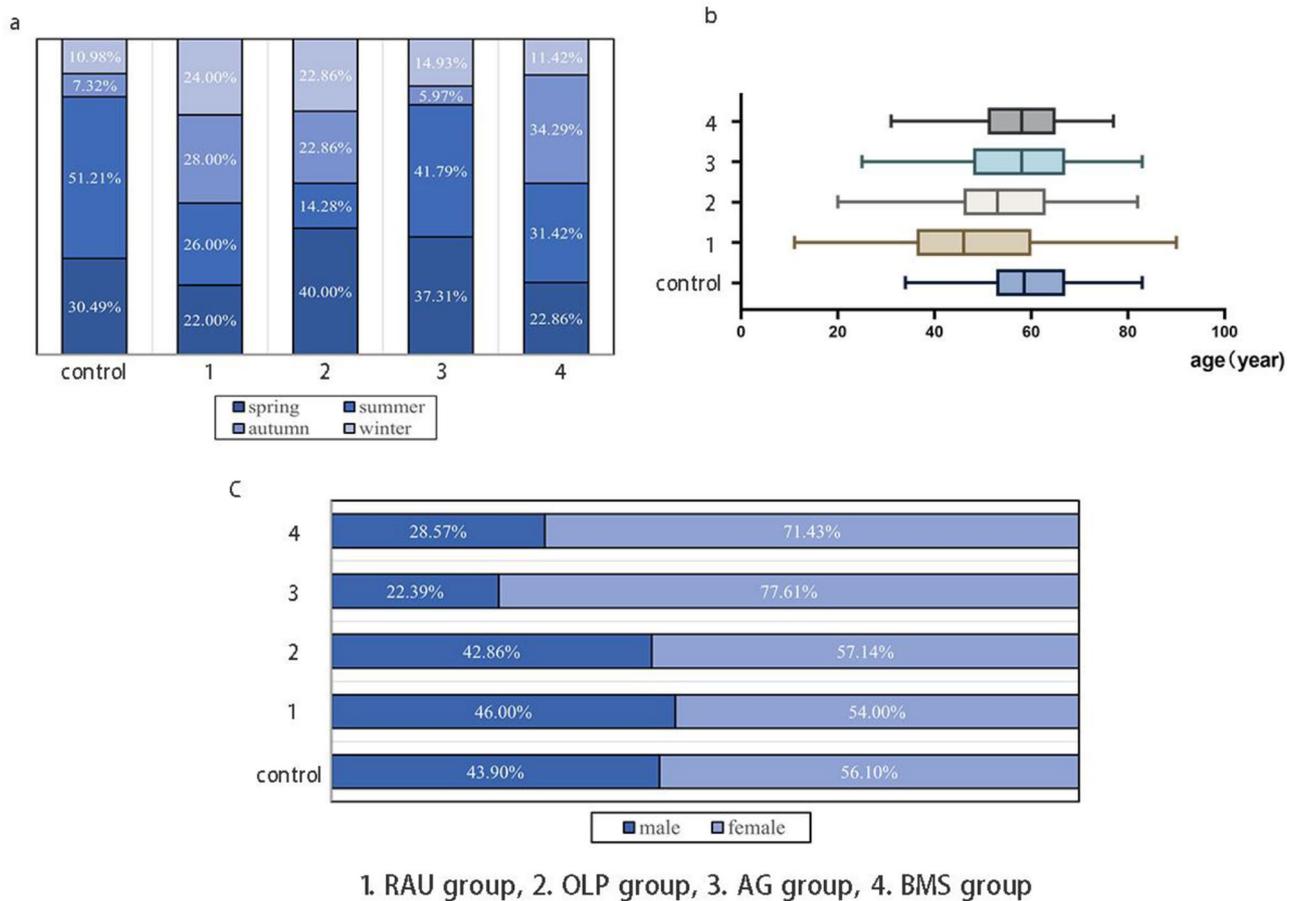


Fig. 2 Demographic distribution of the patients on different oral mucosal lesions. The onset season (A), age (B) and gender (C) of patients with different diseases

constrained, allowing growth until the minimum node size was reached, balancing model complexity and generalization. The model's accuracy, Kappa statistics, and confusion matrices were used to assess its performance, while the importance of individual features was evaluated using the Mean Decrease Accuracy and Gini Index.

Results

General information of the subjects studied

Significant differences were observed in sex, age, and onset season among the healthy control group, recurrent aphthous ulcers (RAU), oral lichen planus (OLP), atrophic glossitis (AG), and burning mouth syndrome (BMS) groups (* $p < 0.05$, Table 1; Fig. 2). Figure 2

provided a demographic analysis of patients across the five groups. Figure 2A showed the seasonal distribution of patients within each group. The AG group was more likely to occur in autumn and winter, while the RAU group and the OLP group were distributed evenly in seasons. The BMS group had the highest percentage in winter (34.29%). Figure 2B displayed the age distribution of patients in each group. The RAU group had the widest range of onset age and the youngest median age of 40, while others preferred to occur in middle age and elderly. Figure 2C presented the gender distribution, where the AG group and the BMS group revealed high incidence in females, reaching 77.61% and 71.43, respectively, while females were only slightly higher than males in the RAU group and the OLP group. The demographic information indicated that oral mucosal diseases' prevalences were related to seasonality, age, and gender.

Comparative analysis of peripheral blood biomarkers in oral mucosal lesion groups

Comparison of peripheral blood humoral immunity indexes in each group

The four oral mucosal lesion groups were elevated compared with the control group, containing complement proteins (C3), complement proteins (C4), IgG, IgM, and IgA, but there was no statistically significant difference ($*p > 0.05$, Table 2).

Comparison of peripheral blood vitamin B levels in each group

The analysis revealed that the levels of vitamin B2 (VB2) and vitamin B3 (VB3) in patients with oral mucosal lesions were significantly higher than those in the control group, and the RAU group exhibiting the highest vitamin

B levels overall. Conversely, vitamin B5 (VB5) levels were significantly lower in the patient groups compared to the control group ($*p < 0.05$, Table 2; Fig. 3A-C), while no significant difference was found in the levels of vitamin B1 ($*p > 0.05$, Table 2). Figure 3 showed that the RAU group had the widest range for VB2 and VB3, with values exceeding 70 and 100, respectively. For VB5, the OLP and RAU groups displayed considerable variability, with some values reaching up to 150, indicating a broader dispersion in these groups.

Comparison of peripheral blood micronutrient levels in each group

Statistically significant differences were observed in serum zinc (Serum Zn), total iron-binding capacity (TIBC), and iron saturation (Iron Sat) levels between the four patient groups with oral mucosal lesions and the healthy control group ($*p < 0.05$, Table 2; Fig. 3D-G), while no significant differences were found in UIBC levels ($*p > 0.05$). Figure 3 illustrated that Serum Zn levels were relatively stable among all groups, mostly ranging from 10 to 20. Serum iron (Serum Fe) levels showed greater variability, particularly in the AG groups, with some values reaching up to 40. TIBC demonstrated substantial variability in the OLP and RAU groups, with some measurements in the RAU group exceeding 80. Iron Sat showed considerable overlap between the control and RAU groups, generally ranging from 20 to 50.

Random forest model performance and biomarker analysis for oral mucosal disease classification

Model performance

We utilized the Random Forest model to assess the performance of different biomarkers. This model achieved

Table 2 Comparison of peripheral blood biomarker levels in different oral mucosal lesions

indexes	health(n=82)	RAU(n=100)	OLP(n=35)	AG(n=67)	BMS(n=35)	F	P
humoral immune indexes	IgG(g/L)	12.07 ± 2.48 ^b	14.82 ± 14.37 ^a	12.67 ± 2.30	12.51 ± 2.50	12.53 ± 1.73	1.536 0.191
	IgA(g/L)	2.44 ± 0.94	2.63 ± 1.07	2.57 ± 0.99	2.35 ± 0.89	2.49 ± 0.79	1.008 0.403
	IgM(g/L)	1.06 ± 0.44	1.17 ± 0.56	1.15 ± 0.48	1.11 ± 0.52	1.03 ± 0.47	0.757 0.554
	C3(g/L)	1.18 ± 0.21	1.16 ± 0.22	1.16 ± 0.19	1.14 ± 0.173	1.16 ± 0.24	0.383 0.821
	C4(g/L)	0.28 ± 0.085 ^d	0.30 ± 0.10	0.28 ± 0.10	0.31 ± 0.11 ^{ae}	0.27 ± 0.08 ^d	2.042 0.088
vitamin B content	VB1(ng/ml)	2.13 ± 1.35	2.09 ± 1.25	2.10 ± 1.18	2.68 ± 3.00	2.06 ± 1.23	1.412 0.23
	VB2(ng/ml)	6.58 ± 4.38 ^{bd}	13.29 ± 16.34 ^a	9.56 ± 7.66	11.47 ± 15.01 ^a	10.54 ± 9.60	3.481 0.008*
	VB3(ng/ml)	18.83 ± 8.70 ^b	26.24 ± 21.28 ^{ace}	19.92 ± 11.01 ^b	23.70 ± 15.31	19.84 ± 13.58 ^b	3.135 0.015*
	VB5(ng/ml)	34.27 ± 19.19 ^{cd}	32.39 ± 19.52 ^c	25.32 ± 9.38 ^{ab}	28.07 ± 10.27 ^a	28.11 ± 9.14	2.972 0.02*
micronutrient levels	Serum Zn(μmol/L)	14.55 ± 1.74 ^{bcde}	12.84 ± 2.70 ^a	13.15 ± 2.60 ^a	13.46 ± 2.21 ^a	12.82 ± 1.85 ^a	7.405 <0.001*
	Serum Fe(μmol/L)	19.62 ± 4.70 ^{bcde}	14.44 ± 5.55 ^a	16.32 ± 6.12 ^a	16.44 ± 7.59 ^a	15.37 ± 4.83 ^a	9.301 <0.001*
	UIBC(μmol/L)	40.48 ± 7.17	42.18 ± 11.95 ^d	38.35 ± 9.66	38.19 ± 10.66 ^b	38.85 ± 9.08	2.102 0.08
	TIBC(μmol/L)	60.10 ± 6.33 ^{bcde}	56.61 ± 9.52 ^a	54.67 ± 7.75 ^a	54.18 ± 7.49 ^a	54.22 ± 7.34 ^a	6.82 <0.001*
	iron saturation(%)	32.84 ± 7.84 ^b	26.37 ± 11.01 ^{ad}	30.17 ± 11.57	30.13 ± 13.44 ^b	28.77 ± 9.83	4.181 0.003*

Note: After the Normality test and homogeneity of variance test, the measurement data of each group were consistent with normal distribution. T-test was used for comparison of the two groups, one-way ANOVA was used for comparison among multiple groups, and LDS test was used for comparison between two groups. ^a $P < 0.05$ compared with the healthy group, ^b $P < 0.05$ compared with the RAU group, ^c $P < 0.05$ compared with the OLP group, ^d $P < 0.05$ compared with the AG group, ^e $P < 0.05$ compared with the BMS group, ^{*} $p < 0.05$

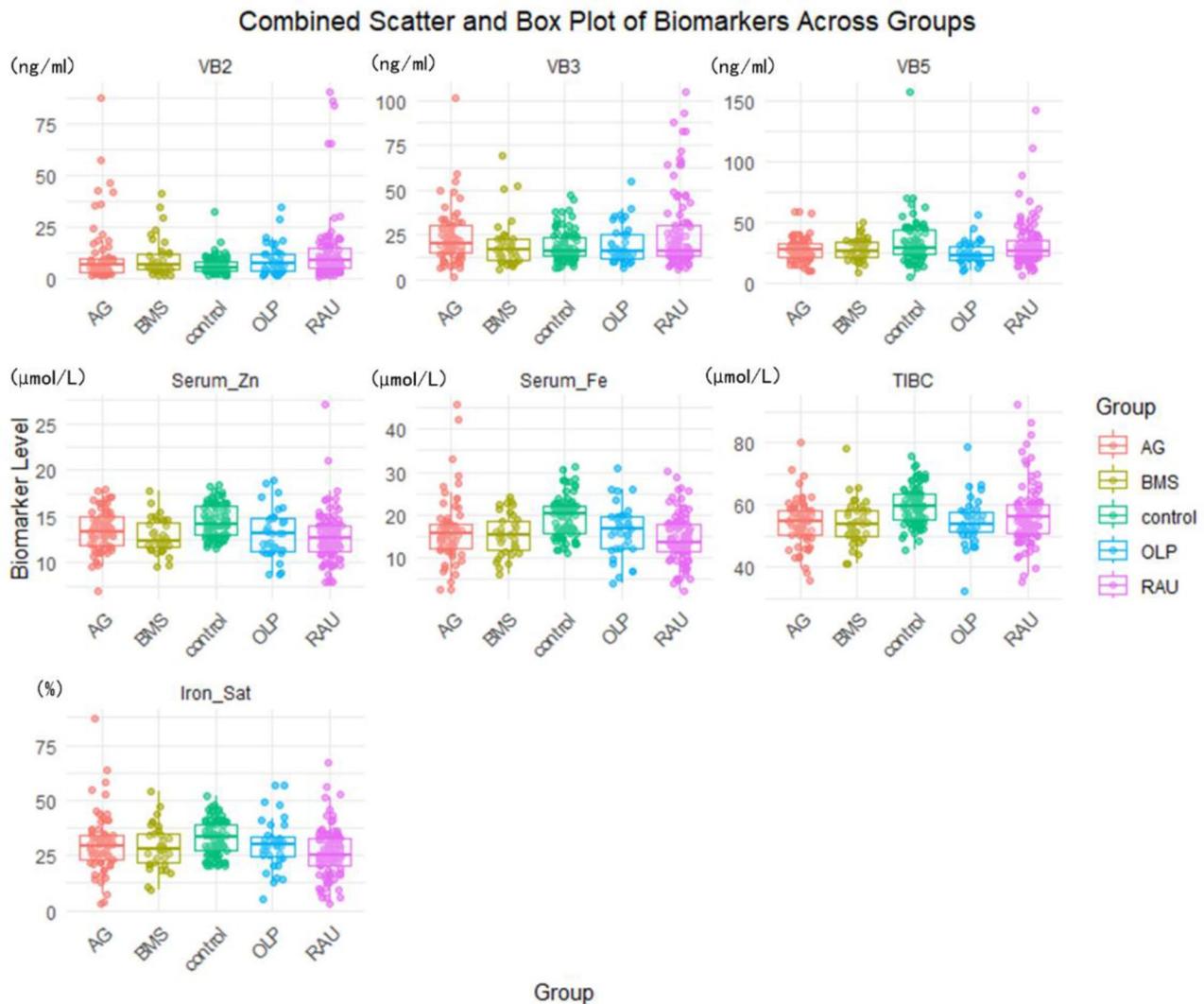


Fig. 3 Differential peripheral blood biomarker levels in different oral mucosal lesions. The different levels of VB2 (A), VB3 (B), VB5 (C), Serum Zn (D), Serum Fe (E), TIBC (F), and Iron Sat (G) within different oral mucosal diseases

Table 3 Sensitivity and specificity of random forest model on different oral mucosal diseases

Actual/Predicted	AG	BMS	Control	OLP	RAU	Class Error
AG	46	1	0	0	0	0.021
BMS	7	18	0	0	0	0.280
Control	0	0	58	0	0	0.000
OLP	3	0	0	22	0	0.120
RAU	0	0	0	0	70	0.000

an accuracy of 94.68% with a Kappa statistic of 0.9306, indicating substantial agreement beyond chance, and revealed great capabilities in classifying different oral mucosal lesions. All RAU cases were identified as RAU, as was the control group. Only 1 AG case was classified as BMS, with a Class Error of 0.021. In the OLP group, 3 cases were considered AG, and the Class Error was 0.120. Unfortunately, the model misclassified 7 BMS cases as AG, resulting in a 28% error rate (Table 3).

Furthermore, we evaluated sensitivity and specificity (Table 4), drew different receiver operating characteristic (ROC) curves according to which (Fig. 4), and calculated the Area Under the Curve (AUC) of the Random Forest model on different oral mucosal diseases (Table 4). Sensitivity measures the model’s ability to correctly identify true positive cases, while specificity indicates its accuracy in identifying true negatives. The model performed well in RAU, AG, and the control group with high percentages

Table 4 Sensitivity, specificity, and AUC for random forest model on different oral mucosal diseases

Group	Sensitivity (%)	Specificity (%)	AUC
Recurrent Aphthous Ulcers (RAU)	96.7	96.88	0.88
Burning Mouth Syndrome (BMS)	40.0	100.0	0.70
Control Group	100.0	98.57	0.99
Oral Lichen Planus (OLP)	70.0	98.81	0.84
Atrophic Glossitis (AG)	95.0	90.54	0.93
Overall AUC			0.8875

of all parameters, while BMS only got 0.7 of AUC for its misclassifying. The overall AUC for the model was 0.8875, indicating strong distinguishing power in differentiating the disease groups.

The importance of biomarkers in classifying diseases

The model identified several key biomarkers contributing significantly to disease classification, including VB2, VB3, Serum Fe, TIBC, and Serum Zn. Figure 5A depicts the accuracy of each biomarker affecting the model, while Fig. 5B indicates the importance of corresponding features in classifying diseases. The Mean Decrease

Accuracy is an indicator for measuring the importance of variables, representing the degree of decrease in model prediction accuracy when the value of a variable is randomly shuffled in a random forest. It revealed that Serum Fe played the most significant role in prediction, followed by VB2, TIBC, Serum Zn, VB3, Iron Sat, and VB5 (Fig. 5A). The Mean Decrease Gini index was a metric indicating how each variable contributes to the homogeneity of the nodes and leaves in the random forest model, and the higher of the index, the more significant of the biomarker. Specifically, VB3 has the highest Mean Decrease Gini score of approximately 23.5, followed closely by VB2 with a score of around 22. Serum Iron (Serum Fe) and TIBC also show significant importance with scores of about 21 and 19.5, respectively. Serum Zinc (Serum Zn) and Vitamin B5 (VB5) have moderate importance with scores near 18, while Iron Saturation (Iron Sat) has the lowest score of approximately 17 (Fig. 4B). These scores highlight the varying impact of each biomarker on the model’s performance, with VB3 and VB2 contributing the most to disease classification.

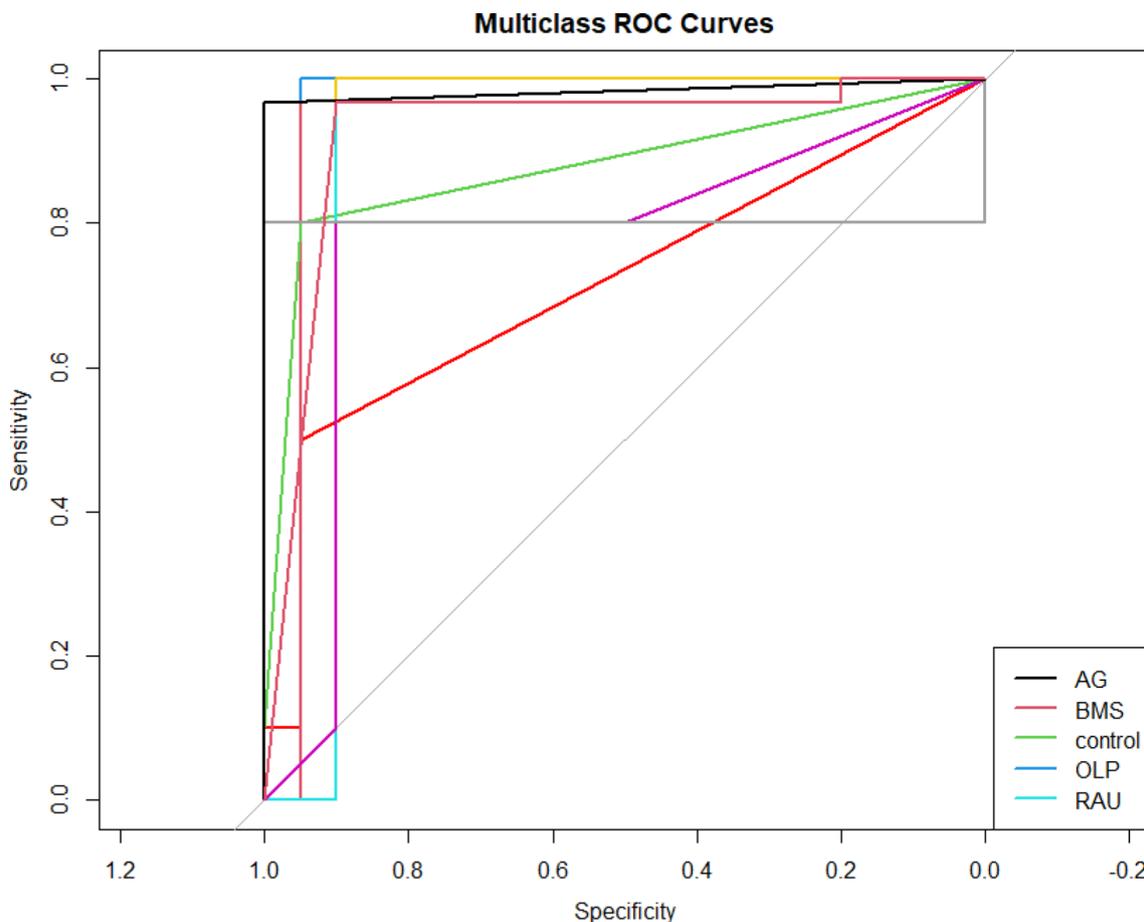


Fig. 4 Sensitivity, specificity, and corresponding multiclass ROC curves for oral mucosal disease classification using Random Forest model

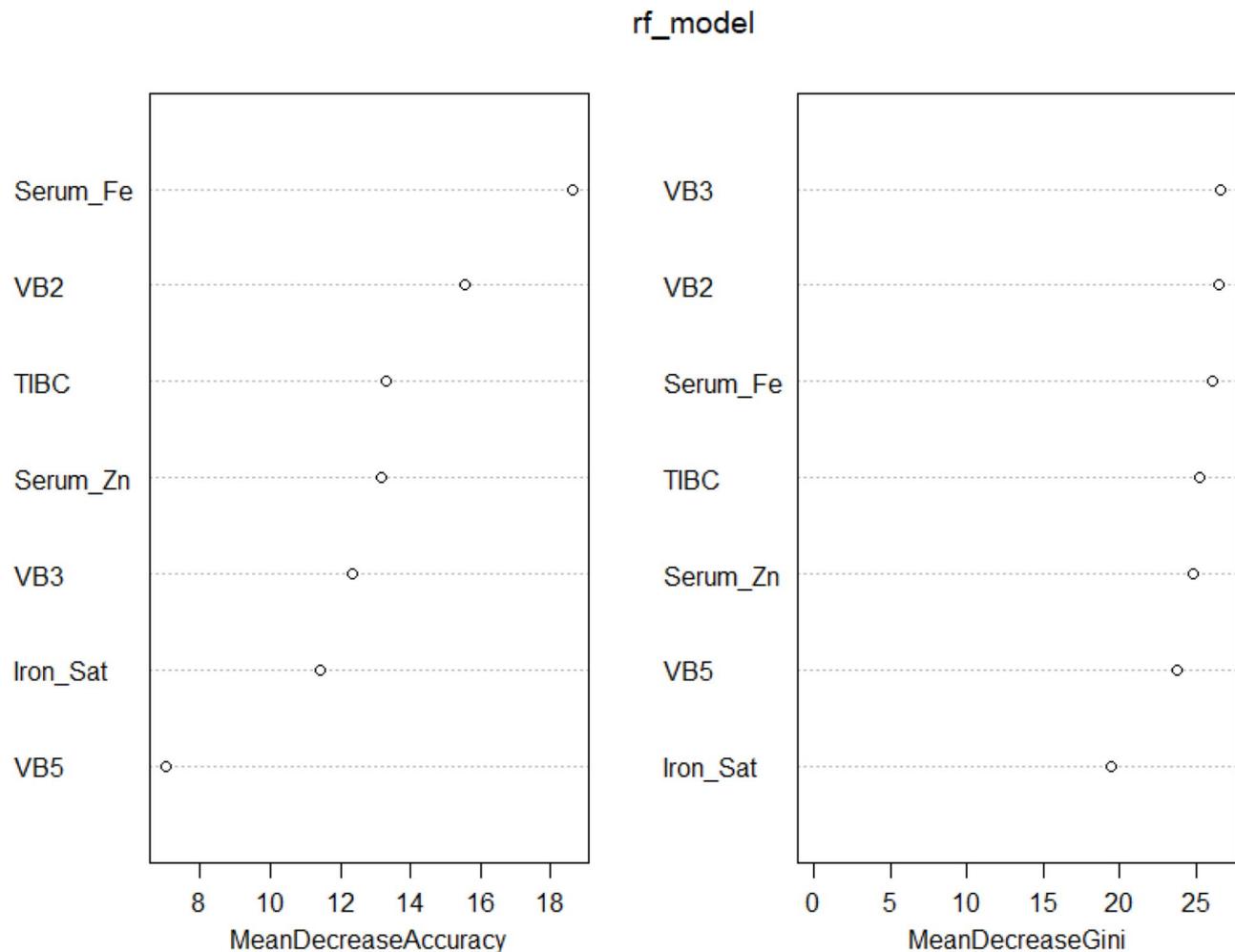


Fig. 5 The heatmap of differential biomarker levels on oral mucosal diseases

Comparative analysis of biomarkers

A comparative analysis of peripheral blood biomarkers across different patient groups revealed distinct patterns. Elevated levels of VB2 and VB3 were observed predominantly in the RAU and BMS groups, while reduced levels of VB5, Serum Fe, and Iron Sat were noted in the OLP and AG groups, as visualized in Fig. 2. Figure 5 visually represents the average levels of various biomarkers (VB2, VB3, VB5, Serum Zn, Serum Fe, TIBC, Iron Sat) across different patient groups (RAU, OLP, control, BMS, AG). The color gradient indicates the intensity of each biomarker, with darker shades representing higher average levels. Notable patterns include the highest level of TIBC in the control group (60.1), compared to 56.61 in the RAU group, 54.67 in the OLP group, 54.22 in the BMS group, and 54.18 in the AG group. The VB2 levels are elevated in the RAU group (13.29), compared to lower levels in the control group (6.58). Similarly, VB3 levels are highest in the RAU group (26.23) and AG group (23.7) while the control group shows a lower average level (18.83). Iron Saturation is relatively consistent but slightly

higher in the control group (32.84) compared to other groups like AG (30.13) and BMS (28.77). These variations underscore the differential biomarker profiles among the groups, which are pivotal for improving the model's diagnostic precision and understanding of the underlying pathophysiology of these diseases (Fig 6).

Discussion

Recurrent aphthous ulcers (RAU), oral lichen planus (OLP), atrophic glossitis (AG), and burning mouth syndrome (BMS) were common oral mucosal lesions, which were related to systemic conditions, such as the deficiency of vitamins. Despite having various therapeutic regimens, palliative treatment was still mainly applied, aiming at pain relief rather than disease cure [19, 20]. Previous attempts to correlate oral mucosal lesions with systemic deficiencies, particularly vitamin B12, folic acid, and iron, have yielded inconclusive results, with some studies finding associations while others not [21–23]. Machine learning (ML) has gained prominence in healthcare for its ability to classify and predict diseases, which

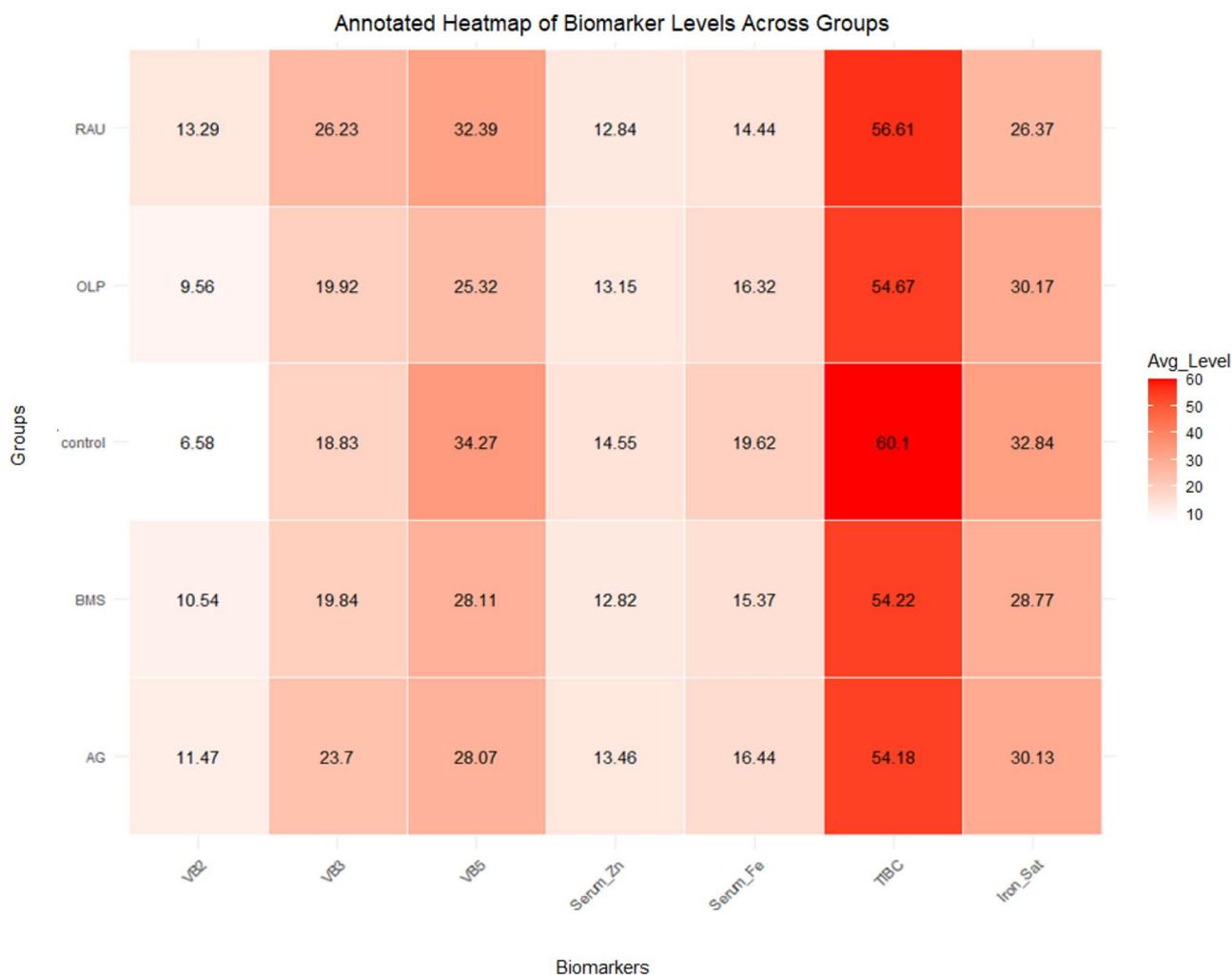


Fig. 6 Annotated heatmap of biomarker levels across group

could handle high-dimensional data, avoid overfitting, and provide interpretability through feature importance metrics [15, 24]. Therefore, it was a reliable method to distinguish whether biomarkers were effective enough in the clinical diagnosis and elucidate the etiology of these diseases. In the context of oral mucosal diseases, ML offers a non-invasive diagnostic tool by analyzing biomarkers from peripheral blood samples [25].

This study investigated the role of micronutrients, humoral immune indexes, and serum vitamin levels in the pathogenesis of oral mucosal lesions. Significant differences ($*p < 0.05$) were found between patients with oral mucosal lesions and healthy controls, especially in serum levels of Vitamin B2 (VB2), vitamin B3 (VB3), vitamin B5 (VB5), serum zinc (Serum Zn), iron (Serum Fe), and iron-binding capacities, which supported the hypothesis that deficiencies or dysregulations in key micronutrients may play a pivotal role in the development of oral mucosal diseases. ANOVA analysis identified key biomarkers that significantly differed between patient groups and

controls, ensuring that only the most relevant biomarkers were included in the subsequent machine-learning analysis to enhance the efficiency and interpretability of the Random Forest model. The application of the Random Forest model further validated the findings from the ANOVA analysis. The model achieved a high classification accuracy of 94.68%, with a Kappa statistic of 0.9306, indicating substantial agreement beyond chance. These metrics suggested that the selected biomarkers, particularly VB2, VB3, Serum Zn, Serum Fe, and total iron-binding capacity (TIBC), were not only statistically significant but also highly predictive of oral mucosal diseases.

Immunological factors included aberrant humoral immune function, aberrant cellular immune function, and autoimmunity. Commonly employed indicators of humoral immune analysis include complement C3, complement C4, antinuclear antibodies, immunoglobulins, and markers of inflammatory response. The most commonly employed method for evaluating humoral immune function was measuring serum immunoglobulin, which

could provide insight into the presence of infections, persistent inflammation, immunodeficiency, and systemic immune diseases [26, 27]. The complement proteins C3 and C4 are primarily responsible for neutralizing viruses, promoting phagocytosis, complement activation, and preventing immune complex deposition [28]. The results of this study demonstrated that four oral mucosal disease groups exhibited elevated levels compared to the control group (C3, C4, IgG, IgM, IgA) without statistically significant difference. It may indicate that oral mucosal disease was not strongly related to humoral immunity. Besides, it may also be related to several factors, including the small sample size, the single test index, the lack of valid pairing, the lack of follow-up, and the lack of multi-center study design employed in this study. In the future, it would be beneficial to increase the sample size or analyze the data in conjunction with cellular immunity indicators, among other factors.

Micronutrient levels, particularly vitamins and trace elements, played crucial roles in various physiological processes, whose imbalance could lead to disease [29]. For example, Vitamin B, a water-soluble vitamin that could not be synthesized by the body, must be supplemented externally [30]. Stress, poor diet, certain medications, and high consumption rates could lead to a deficiency of it [31]. Several studies have linked vitamin B to epithelial disease pathogenesis, including its impact on one-carbon metabolism and the risk of conditions like gastric cancer [32]. Vitamin and micronutrient supplementation, such as Vitamin B and Serum Zn, has shown potential in alleviating symptoms of oral mucosal diseases, including pain and burning sensations in burning mouth syndrome [33]. Previous studies have focused on vitamin B12 and lipids in the context of oral mucosal lesions and concluded that vitamin B12 deficiency may be an etiological factor in recurrent aphthous stomatitis (RAS) [34]. However, there has been few investigations of the other subclasses of vitamin B. This study demonstrated that there were significant differences in VB2, VB3, VB5, Serum Zn, Serum Fe, TIBC, and iron Sat among four groups. The levels of VB2 and VB3 were significantly higher than those in the control group ($*p < 0.05$). It has been established that VB2, VB3, and VB5 played essential roles in oxidative reactions and the maintenance of mitochondrial functions [35], which was also vital for the citric acid cycle and its intermediate product [36].

Our study revealed significant differences in VB2, VB3, and VB5 levels between patients with oral mucosal lesions and healthy controls, with VB2 and VB3 being significantly higher in the disease groups ($*p < 0.05$). It suggested a potential dysfunction in the mechanisms regulating oxidative stress. FMN and FAD, were active forms of VB2, known as essential coenzymes for redox enzymes, which played a critical role in the regulation of

oxidative stress by facilitating key metabolic processes, including the oxidative respiratory chain, fatty acid and amino acid oxidation, and the citric acid cycle, which collectively helped maintain redox homeostasis in the body [37]. Another potential mechanism involves the immune regulatory functions of MAIT cells (mucosal-associated invariant T cells), where Vitamin B metabolites play a crucial role in their activation. MAIT cells, abundantly existing on oral mucosal surfaces, were closely linked to Vitamin B2 metabolites [38]. These metabolites circulated throughout the body, influencing the development and maturation of MAIT cells in the thymus, as well as their functional activity and ability to recognize VB2 metabolites [39]. It suggested that VB2 may be immunologically involved in the pathogenesis of oral mucosal diseases. In addition, we observed significantly lower levels of VB5, Serum Zn, Serum Fe, iron-binding capacity, and transferrin saturation in patients with oral mucosal lesions, implicating these micronutrients took effect in the etiology of oral mucosal lesions. Studies have shown an association between oral mucosal diseases and gastrointestinal mucosal disorders [40]. Therefore, it was plausible that gastrointestinal diseases, which would impair the absorption of key micronutrients such as vitamins and minerals, contributing to nutrient deficiencies [41], including vitamins, folate, iron, and zinc, suggesting that the mechanism linking oral mucosal diseases to nutrient deficiencies may be mediated through gastrointestinal disease pathways. For instance, a Zinc-deficient diet could lead to parakeratosis in normally orthokeratinized oral mucosa, while iron participating in collagen synthesis suggested that significant reductions in Serum Fe levels could impair oral mucosal formation. Deficiencies in these elements may result in clinical symptoms such as loss of filiform papillae and abnormal mucosal thickening, which are aligned with the typical manifestations of some oral mucosal diseases [42]. Furthermore, inflammatory changes in the tongue mucosa are often associated with a burning sensation, especially in cases of vitamin B1 deficiency. It can be attributed to the essential role of thiamine (VB1) in energy production, nerve impulse transmission, and the maintenance of the myelin sheath [43]. Previous research has identified correlations between these biomarkers and conditions such as anemia, vitamin B12, and folic acid deficiencies, which are more common in patients with oral mucosal lesions compared to controls [3, 44]. By excluding patients with systemic diseases, this study specifically focused on comparing the differences in these indicators among patients with mucosal diseases, as these parameters could be easily influenced by gender, age, and overall health conditions. It was reported that notable differences in hematological parameters were observed among RAS patients when stratified by gender and age groups, which was in

accordance with our study [45]. The prevalence of serum ferritin deficiency was significantly higher in young and middle-aged female patients [46], while serum folate and vitamin B12 deficiencies were more prevalent in the young adult male population [45]. A study indicated that Cu and Zn concentrations were notably higher in men compared to women. Additionally, it found a positive correlation between body mass index and Cu levels specifically in men, while smoking was linked to reduced Se levels in the male population [47]. It indicated that gender and age could significantly impact hematologic parameters, while other systemic factors, such as smoking and body mass index would also influence these indicators.

Mental factors were also thought related to oral mucosal diseases. It was reported that OLP patients showed higher levels of stress, anxiety and depression, who were also detected higher level of serum cortisol [48]. Nevertheless, there was an opposite viewpoint in RAU. It was reported that anxiety, depression and stress would not increase the risk of RAU, but sleep quality would [49, 50]. Further study is needed to explore the connection between mental variables and oral mucosal lesions.

This study utilized the Random Forest model from machine learning, training and validating with our own dataset. In addition to successfully constructing the training model, it also achieved high accuracy. Machine learning aids in processing complex clinical datasets, enabling automatic identification of biomarkers crucial for disease classification and prediction [51]. For conditions like oral mucosal diseases, which are challenging to diagnose, the application of machine learning in personalized medicine, disease prediction, and treatment monitoring could significantly assist both healthcare providers and patients in health management better [52]. A study reported that multiple models had been successfully constructed, utilizing 50 features derived from blood routine and biochemical detection data, to diagnose various cardiovascular diseases [53]. With larger datasets and more rigorous study designs, machine learning algorithms could lead to the development of highly accurate disease prediction models, which could be stably applied in clinical practice at the personalized level in the future [54].

This study examines a range of common oral mucosal lesions, comparing them with healthy controls to provide insights into the biomarkers associated with these conditions. By analyzing both traditional immune parameters (e.g., immunoglobulins and complements) and micronutrient levels, the study offered a comprehensive understanding of their potential role in the pathophysiology of oral mucosal lesions. The model demonstrated high diagnostic precision in classifying these diseases, with feature importance analysis highlighting key biomarkers,

including VB2, VB3, Serum Fe, TIBC, and Serum Zn as significant contributors to classification accuracy.

However, there were still several limitations of the Random Forest model, even though it exhibited robust performance. First, despite the model's overall high accuracy, it showed lower sensitivity in classifying BMS cases, indicating a more complex etiology of this condition and potential data imbalances in the dataset. It was a common issue in machine learning models, where minority classes were often underrepresented, leading to poorer classification performance for these groups. To address this problem, future studies should balance datasets more effectively, either through over-sampling, under-sampling, or advanced methods such as SMOTE (Synthetic Minority Over-sampling Technique) [55, 56]. Moreover, the cross-sectional design of this study limited the ability to draw causal inferences between micronutrient levels and disease onset or progression. Longitudinal studies were needed to clarify whether these biomarkers played roles in disease initiation or were merely a consequence of the disease state. Additionally, while the Random Forest model demonstrated effectiveness in classifying the disease groups, the lack of external validation limited the generalizability of these findings. Despite being robust and widely utilized in clinical research, the Random Forest model had certain limitations that need to be addressed, one of which was the risk of overfitting, particularly when the sample size was small [57]. Meanwhile, the model's effectiveness in clinical applications largely depends on its ability to generalize across diverse patient populations. Factors such as demographic variability, regional differences in disease prevalence, and variations in clinical settings could significantly impact its generalizability [58, 59]. Therefore, Future research should prioritize validation of the Random Forest model in larger, multicenter cohorts to confirm its generalizability across diverse populations. Additionally, building upon the biomarker associations identified in this study, future investigations could employ correlation analysis, multi-omics integration, or causal modeling approaches to dissect the interplay between micronutrient profiles. Another limitation lied in the scope of biomarkers analyzed. While this study focused on micronutrients and selected humoral immune indexes, the inclusion of cellular immune markers and other relevant biomarkers, such as inflammatory cytokines, fat-soluble vitamins (A, D, E, K), other water-soluble vitamins like vitamin C, and clinical phenotypes such as age, sex, pain symptoms, grade of the lesion could provide a more comprehensive understanding of the pathophysiology underlying oral mucosal lesions [60]. Especially, due to the overlap in clinical presentations of different mucosal diseases, further rigorous research considering disease severity was essential to refine the study's conclusions. For example, RAS ranges from mild

to severe, and oral lichen planus includes both typical and severe ulcerative or bullous types. Whether variations in the variables included in this study existed between different subtypes of the same disease, or whether similar clinical symptoms, such as ulcers, lead to differing conclusions for different diseases, required further investigation. Furthermore, exploring other machine learning algorithms, such as Support Vector Machines (SVM) [61] or Neural Networks [62], could enhance diagnostic accuracy, particularly for diseases like BMS, where classification remains challenging.

Conclusion

This study demonstrated the utility of combining traditional statistical methods, such as ANOVA, with machine learning techniques like Random Forest to classify oral mucosal diseases. Key biomarkers, such as vitamins B2, B3, B5, serum iron, and zinc, were identified as significantly associated with these conditions. The Random Forest model achieved high classification accuracy (94.68%), highlighting its potential to improve diagnostic precision. These findings suggested that monitoring the levels of specific vitamins and micronutrients could aid in the diagnosis and management of oral mucosal lesions. However, further research with larger cohorts and external validation is needed to confirm these associations and optimize the application of machine learning in clinical diagnostics.

Abbreviations

RAU	Recurrent aphthous ulcers
OLP	Oral lichen planus
AG	Atrophic glossitis
BMS	Burning mouth syndrome
IgG	Immunoglobulin G
IgA	Immunoglobulin A
IgM	Immunoglobulin M
C3	Complement C3
C4	Complement C4
VB1	Vitamin B1
VB2	Vitamin B2
VB3	Vitamin B3
VB5	Vitamin B5
UIBC	Unsaturated iron-binding capacity
TIBC	Total iron-binding capacity
ML	Machine learning
RMF	Random Forest Model

Acknowledgements

We thank all the volunteers who participated in this study. Thanks to Zixin Zhou for the assistance provided during the revision process.

Author contributions

Huiyu Yao contributed to the data curation, formal analysis, investigation, validation, visualization, and the writing, review & editing of the manuscript; Zixin Cao, Liangfu Huang, Haojie Pan, Xiaomin Xu contributed to the data curation; Fucui Sun contributed to the writing, review & editing of the manuscript; Xi Ding and Wan Wu contributed to the conceptualization, data curation, funding acquisition, methodology, project administration, supervision and the review & editing of the manuscript; All authors approved the submission of the final versions of the manuscript. Corresponding author: Wan Wu and Xi Ding were contributed equally to this manuscript.

Funding

This work was supported by the Basic Scientific Research Project of Wenzhou (Y2023289).

Data availability

The datasets collected and analyzed during the current study are not publicly available due to patient privacy restrictions under the General Data Protection Regulation (GDPR). Anonymized data are available from the corresponding author upon reasonable request, pending approval from the Institutional Review Board (IRB) of The First Affiliated Hospital of Wenzhou Medical University and a signed data use agreement.

Declarations

Ethics approval and consent to participate

All data were collected from patients who gave their informed consent themselves or through legal representatives. This study is a retrospective analysis utilizing existing clinical data. The research protocol has been approved by Ethics Committee in Clinical Research (ECCR) of the First Affiliated Hospital of Wenzhou Medical University (No. KY2024-R233), and the conduct of the study adheres to the Declaration of Helsinki and relevant ethical standards. Due to the nature of the study, patient data involved have been de-identified, and thus the Ethics Committee has waived the requirement for patient informed consent.

Human ethics and consent to participate

This study was approved by the Ethics Committee in Clinical Research (ECCR) of the First Affiliated Hospital of Wenzhou Medical University (Approval No. KY2024-R233). The study was conducted in accordance with the Declaration of Helsinki and its later amendments or comparable ethical standards. For this retrospective analysis, which involved the use of existing clinical data, the Ethics Committee waived the requirement for patient informed consent due to the de-identified nature of the data. However, it is important to note that all patients at our hospital provide general consent for the use of their medical records for teaching and research purposes as part of the informed consent process prior to receiving oral treatment.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Stomatology, The First Affiliated Hospital of Wenzhou Medical University, Nanbaixiang Ouhai District, Wenzhou, Zhejiang 325000, People's Republic of China

Received: 15 October 2024 / Accepted: 2 May 2025

Published online: 10 May 2025

References

1. Wang X. The fourth Chinese oral health epidemiological survey report on the oral health status of Chinese residents. People's Medical Publishing House; 2018.
2. Maymone MBC, Greer RO, Burdine LK, Dao-Cheng A, Venkatesh S, Sahitya PC, Maymone AC, Kesecker J, Vashi NA. Benign oral mucosal lesions: clinical and pathological findings. *J Am Acad Dermatol*. 2019;81(1):43–56.
3. Mousavi T, Jalali H, Moosazadeh M. Hematological parameters in patients with recurrent aphthous stomatitis: a systematic review and meta-analysis. *BMC Oral Health*. 2024;24(1):339.
4. Stoopler ET, Villa A, Bindakhil M, Díaz DLO, Sollecito TP. Common oral conditions: A review. *JAMA*. 2024;331(12):1045–54.
5. de Lanna CA, da Silva BNM, de Melo AC, Bonamino MH, Alves LDB, Pinto LFR, Cardoso AS, Antunes HS, Boroni M, Cohen Goldemberg D. oral lichen planus and oral squamous cell carcinoma share key oncogenic signatures. *Sci Rep*. 2022;12(1):20645.

6. Li H, Sun J, Wang X, Shi J. Oral microbial diversity analysis among atrophic glossitis patients and healthy individuals. *J Oral Microbiol*. 2021;13(1):1984063.
7. Randall DA, Wilson Westmark NL, Neville BW. Common oral lesions. *Am Fam Physician*. 2022;105(4):369–76.
8. Momin S. Burning mouth Syndrome-A frustrating problem. *JAMA Otolaryngol Head Neck Surg*. 2021;147(6):580.
9. Deng X, Wang Y, Jiang L, Li J, Chen Q. Updates on immunological mechanistic insights and targeting of the oral lichen planus microenvironment. *Front Immunol*. 2022;13:1023213.
10. Sun A, Lin H-P, Wang Y-P, Chiang C-P. Significant association of deficiency of hemoglobin, iron and vitamin B12, high homocysteine level, and gastric parietal cell antibody positivity with atrophic glossitis. *J Oral Pathol Med*. 2012;41(6):500–4.
11. Chang JY-F, Wang Y-P, Wu Y-C, Cheng S-J, Chen H-M, Sun A. Blood profile of oral mucosal disease patients with both vitamin B12 and iron deficiencies. *J Formos Med Assoc*. 2015;114(6):532–8.
12. Şenel S. An overview of physical, microbiological and immune barriers of oral mucosa. *Int J Mol Sci* 2021, 22(15).
13. Chiang C-P, Yu-Fong Chang J, Wang Y-P, Wu Y-H, Wu Y-C, Sun A. Recurrent aphthous stomatitis - Etiology, serum autoantibodies, anemia, hematinic deficiencies, and management. *J Formos Med Assoc*. 2019;118(9):1279–89.
14. Wu Y-C, Wu Y-H, Wang Y-P, Chang JY-F, Chen H-M, Sun A. Hematinic deficiencies and anemia statuses in recurrent aphthous stomatitis patients with or without atrophic glossitis. *J Formos Med Assoc*. 2016;115(12):1061–8.
15. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol*. 2022;23(1):40–55.
16. Machoy ME, Szyszka-Sommerfeld L, Vegh A, Gedrange T, Woźniak K. The ways of using machine learning in dentistry. *Adv Clin Exp Med*. 2020;29(3):375–84.
17. International Classification of Diseases. (ICD-11) [<https://icd.who.int/>]
18. Peterson DE, Boers-Doets CB, Bensadoun RJ, Herrstedt J. Management of oral and Gastrointestinal mucosal injury: ESMO clinical practice guidelines for diagnosis, treatment, and follow-up. *Ann Oncol*. 2015;26(Suppl 5):v139–51.
19. Chung M-K, Wang S, Oh S-L, Kim YS. Acute and chronic pain from facial skin and oral mucosa: unique neurobiology and challenging treatment. *Int J Mol Sci* 2021, 22(11).
20. Sankar V, Hearnden V, Hull K, Juras DV, Greenberg MS, Kerr AR, Lockhart PB, Patton LL, Porter S, Thornhill M. Local drug delivery for oral mucosal diseases: challenges and opportunities. *Oral Dis*. 2011;17(Suppl 1):73–84.
21. Kozlak ST, Walsh SJ, Lalla RV. Reduced dietary intake of vitamin B12 and folate in patients with recurrent aphthous stomatitis. *J Oral Pathol Med*. 2010;39(5):420–3.
22. Lopez-Jornet P, Camacho-Alonso F, Martos N. Hematological study of patients with aphthous stomatitis. *Int J Dermatol*. 2014;53(2):159–63.
23. Sun A, Chen H-M, Cheng S-J, Wang Y-P, Chang JY-F, Wu Y-C, Chiang C-P. Significant association of deficiencies of hemoglobin, iron, vitamin B12, and folic acid and high homocysteine level with recurrent aphthous stomatitis. *J Oral Pathol Med*. 2015;44(4):300–5.
24. Adeoye J, Zheng L-W, Thomson P, Choi S-W, Su Y-X. Explainable ensemble learning model improves identification of candidates for oral cancer screening. *Oral Oncol*. 2023;136:106278.
25. Zhu X, Wang C-L, Yu J-F, Weng J, Han B, Liu Y, Tang X, Pan B. Identification of immune-related biomarkers in peripheral blood of schizophrenia using bioinformatic methods and machine learning algorithms. *Front Cell Neurosci*. 2023;17:1256184.
26. Isabel FFM, Gisel V, Brett U, Julio M, CM HA. Higher serum levels of systemic inflammatory markers are linked to greater inspiratory muscle dysfunction in COPD. *Clin Respir J*. 2019;13(4):247–55.
27. Bunker JJ, Bendelac A. IgA responses to microbiota. *Immunity*. 2018;49(2):211–24.
28. Li H, Lin S, Yang S, Chen L, Zheng X. Diagnostic value of serum complement C3 and C4 levels in Chinese patients with systemic lupus erythematosus. *Clin Rheumatol*. 2015;34(3):471–7.
29. Peterson CT, Rodionov DA, Osterman AL, Peterson SN. B vitamins and their role in immune regulation and Cancer. *Nutrients* 2020, 12(11).
30. Eckle SBG, Corbett AJ, Keller AN, Chen Z, Godfrey DI, Liu L, Mak JYW, Fairlie DP, Rossjohn J, McCluskey J. Recognition of vitamin B precursors and byproducts by mucosal associated invariant T cells. *J Biol Chem*. 2015;290(51):30204–11.
31. Gille D, Schmid A. Vitamin B12 in meat and dairy products. *Nutr Rev*. 2015;73(2):106–15.
32. Miranti EH, Stolzenberg-Solomon R, Weinstein SJ, Selhub J, Männistö S, Taylor PR, Freedman ND, Albanes D, Abnet CC, Murphy GA-O. Low vitamin B(12) increases risk of gastric cancer: A prospective study of one-carbon metabolism nutrients and risk of upper Gastrointestinal tract cancer. *Int J Cancer*. 2017;141(6):1120–9. (1097–0215 (Electronic)).
33. Jankovskis V, Selga G. Vitamin B and zinc supplements and capsaicin oral rinse treatment options for burning mouth syndrome. *Med (Kaunas)* 2021, 57(4).
34. Piskin S, Sayan C, Durukan N, Senol M. Serum iron, ferritin, folic acid, and vitamin B12 levels in recurrent aphthous stomatitis. *J Eur Acad Dermatol Venereol*. 2002;16(1):66–7.
35. Depeint F, Bruce WR, Shangari N, Mehta R, O'Brien PJ. Mitochondrial function and toxicity: role of the B vitamin family on mitochondrial energy metabolism. *Chem Biol Interact* 2006, 163(1–2).
36. Ford TC, Downey LA, Simpson T, McPhee G, Oliver C, Stough C. The Effect of a High-Dose Vitamin B Multivitamin Supplement on the Relationship between Brain Metabolism and Blood Biomarkers of Oxidative Stress: A Randomized Control Trial. *Nutrients* 2018, 10(12).
37. Doroftei B, Ilie O-D, Cojocariu R-O, Ciobica A, Maftei R, Grab D, Anton E, McKenna J, Dhunna N, Simionescu G. Minireview exploring the biological cycle of vitamin B3 and its influence on oxidative stress: further molecular and clinical aspects. *Molecules* 2020, 25(15).
38. Ioannidis M, Cerundolo V, Salio M. The immune modulating properties of Mucosal-Associated invariant T cells. *Front Immunol*. 2020;11:1556.
39. Legoux F, Salou M, Lantz O. MAIT cell development and functions: the microbial connection. *Immunity*. 2020;53(4):710–23.
40. Kunath BJ, De Rudder C, Laczny CC, Letellier E, Wilmes P. The oral-gut Microbiome axis in health and disease. *Nat Rev Microbiol*. 2024;22(12):791–805.
41. Massironi S, Viganò C, Palermo A, Pirola L, Mulinacci G, Allocca M, Peyrin-Biroulet L, Danese S. Inflammation and malnutrition in inflammatory bowel disease. *Lancet Gastroenterol Hepatol*. 2023;8(6):579–90.
42. Bhattacharya PT, Misra SR, Hussain M. Nutritional Aspects of Essential Trace Elements in Oral Health and Disease: An Extensive Review. *Scientifica (Cairo)* 2016, 2016:5464373.
43. Lešić S, Ivanišević Z, Špiljak B, Tomas M, Šošarić M, Včev A. The impact of vitamin deficiencies on oral manifestations in children. *Dent J (Basel)* 2024, 12(4).
44. Lin H-P, Wu Y-H, Wang Y-P, Wu Y-C, Chang JY-F, Sun A. Anemia and hematinic deficiencies in anti-gastric parietal cell antibody-positive or all autoantibodies-negative recurrent aphthous stomatitis patients. *J Formos Med Assoc* 2017, 116(2).
45. Bao ZX, Shi J, Yang XW, Liu LX. Hematinic deficiencies in patients with recurrent aphthous stomatitis: variations by gender and age. *Med Oral Patol Oral Cir Bucal*. 2018;23(2):e161–7.
46. Mei Z, Addo OY, Jeffers ME, Sharma AJ, Flores-Ayala RC, Brittenham GM. Physiologically based serum ferritin thresholds for iron deficiency in children and non-pregnant women: a US National health and nutrition examination surveys (NHANES) serial cross-sectional study. *Lancet Haematol*. 2021;8(8):e572–82.
47. Benes B, Speváčková V, Smíd J, Batářiová A, Cejchanová M, Zítková L. Effects of age, BMI, smoking and contraception on levels of Cu, se and Zn in the blood of the population in the Czech Republic. *Cent Eur J Public Health*. 2005;13(4):202–7.
48. Khan S, Mehta DN, Jain P, Somani S, Pathan MA, Thakkar H, Agrawal S. A study to assess the role of psychological stress in the severity of oral lichen planus, OSMF, and leukoplakia and its correlation with serum cortisol levels. *J Pharm Bioallied Sci*. 2024;16(Suppl 3):S2021–3.
49. Mirzaei M, Zarabadipour M, Mirzadeh M. Evaluation the relationship between psychological profile and salivary cortisol in patients with recurrent aphthous stomatitis. *Dent Res J (Isfahan)*. 2021;18:50.
50. Gao X, Chen P, Liu J, Fan X, Wu Z, Fang H, Zhang Z. Sleep quality and perceived stress levels in Chinese patients with minor recurrent aphthous stomatitis: a cross-sectional questionnaire-based survey. *Postgrad Med*. 2024;136(7):749–56.
51. Komorowski M, Green A, Tatham KC, Seymour C, Antcliff D. Sepsis biomarkers and diagnostic tools with a focus on machine learning. *EBioMedicine*. 2022;86:104394.
52. Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920–30.
53. Wang Z, Gu Y, Huang L, Liu S, Chen Q, Yang Y, Hong G, Ning W. Construction of machine learning diagnostic models for cardiovascular pan-disease based on blood routine and biochemical detection data. *Cardiovasc Diabetol*. 2024;23(1):351.

54. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Intern Med.* 2018;284(6):603–19.
55. Nakamura M, Kajiwara Y, Otsuka A, Kimura H. LVQ-SMOTE - Learning vector quantization based synthetic minority Over-sampling technique for biomedical data. *BioData Min.* 2013;6(1):16.
56. Naseriparsa M, Al-Shammari A, Sheng M, Zhang Y, Zhou R. RSMOTE: improving classification performance over imbalanced medical datasets. *Health Inf Sci Syst.* 2020;8(1):22.
57. Song L, Langfelder P, Horvath S. Random generalized linear model: a highly accurate and interpretable ensemble predictor. *BMC Bioinformatics.* 2013;14:5.
58. Yang J, Soltan AAS, Clifton DA. Machine learning generalizability across healthcare settings: insights from multi-site COVID-19 screening. *NPJ Digit Med.* 2022;5(1):69.
59. Rajendran S, Pan W, Sabuncu MR, Chen Y, Zhou J, Wang F. Learning across diverse biomedical data modalities and cohorts: challenges and opportunities for innovation. *Patterns (N Y).* 2024;5(2):100913.
60. Alkhadar H, Macluskey M, White S, Ellis I, Gardner A. Comparison of machine learning algorithms for the prediction of five-year survival in oral squamous cell carcinoma. *J Oral Pathol Med.* 2021;50(4):378–84.
61. Chu CS, Lee NP, Adeoye J, Thomson P, Choi S-W. Machine learning and treatment outcome prediction for oral cancer. *J Oral Pathol Med.* 2020;49(10):977–85.
62. Xie F, Xu P, Xi X, Gu X, Zhang P, Wang H, Shen X. Oral mucosal disease recognition based on dynamic self-attention and feature discriminant loss. *Oral Dis.* 2024;30(5):3094–107.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.