PLS Discriminant Analysis for Classification of Caries Level

Katrul Nadia Basri, Mohd Norzaliman Mohd Zain, Zalhan Mohd Yusof, Ahmad Sabirin Zoolfakar, Farinawati Yazid*

Abstract— **Saliva has been investigated as a possible diagnostic medium to detect the presence of dental caries. Salivary alpha-amylase has a direct connection with caries activity, making it a useful host-related factor biomarker. The saliva sample from UKM was used as a sample for the measurement of UV absorption spectroscopy. The UV spectra obtained shows the peak absorption around 280 nm that correlate with the amino acid in the salivary alpha amylase that works as a binding site of bacteria that responsible of caries. The spectra was split into calibration and validation for the development of the PLS model. Autoscale and MSC were used as data preprocessing for the model optimization. The best sensitivity and specifity acquired were 1.00 respectively with the AUC value of 1.00. Therefore, the UV absorption spectroscopy coupled with chemometrics can work as diagnostic tool for detection of dental caries.**

Index Terms—**Caries, chemometrics, UV spectroscopy**

I. INTRODUCTION

Dental caries is a progressive disease due to the eroded of mineral on the enamel and dentin due to the acid excreted from the acidogenic bacteria. 92% of adult in worldwide has experienced the dental caries [1]. Saliva plays a crucial role in the saliva-bacteria interaction that provide the colonization site, mechanism of dental plaque and the protein in saliva act as a substrate for the microbial nutrition [2]. Based on the work reported, the saliva can be a biomarker of the starch hydrolysis that bind to the oral bacteria known as streptococci and carbohydrate digestion [4]. However, there is limited diagnostic value for caries prediction because saliva is not representative of the microbial community at the disease site [3].

This manuscript is submitted on 12 June 2023, revised on 2 September 2023, accepted on 31 October 2023 and published on 30 April 2024.

Katrul Nadia Basri and Ahmad Sabirin Zoolfakar with School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Kuala Lumpur (e-mail: katruln@gmail.com).

Mohd Norzaliman Mohd Zain and Zalhan Md Yusof from Photonics Technology Laboratory, MIMOS Berhad, Technology Park Malaysia, 57000 Kuala Lumpur, Malaysia

Farinawati Yazid with Faculty of Dentistry, Universiti Kebangsaan Malaysia, 50300 Kuala Lumpur,

*Corresponding author
Email address: drfarinawati@ukm.edu.my

The early detection is important to reverse the demineralization process [4] [5]. Early detection tool require monitoring of caries progression to assist in the decision making. Many methods has been used for the detection of early caries lesion such as x-ray imaging [6] and tactile inspection [7]. These method require an expert to handle the instrument and make the decision making [8]. The fluorescence imaging was used to access the numerical grade of caries by defining the spectral intensity within range 565-750 nm [9].

The application of spectroscopy in examine the dental health has been known with the implementation of Raman spectroscopy, infrared spectroscopy and UV spectroscopy. Raman and near infrared spectroscopy was used in study of dental hard tissue and its structure. Nao Miyamoto et. al detect the progress in the dental caries using Raman spectroscopy by visualize the diseased zones and monitoring the full width at half maximum (FWHM) of the Raman graph [10]. Near infrared illumination also worked as teeth hyperspectral imaging to detect the presence of caries because NIR light can penetrate deeper into the teeth enamel [11] [12]. The application of UV absorption spectroscopy in detection of dental caries because natural tooth structure exhibits fluorescence when exposed to UV light [13]. Therefore,in this study the application of UV spectroscopy will be applied on saliva sample.

In spectroscopy, the method will be incorporated with chemometrics to extract the informative property of interest. Yong Wang et.al combined UV-VIS spectroscopy with chemometrics to construct a sensitive spectroscopic biosensor using multivariate curve resolution by alternating least square [14]. The implementation of data preprocessing such as autoscale, standard normal variate and multiplicative scatter correction [14] aid in optimizing the developed model.

In this paper, UV absorption spectroscopy combined with chemometrics will be applied to detect the level severity of dental caries. A few method of preprocessing will be implemented to optimized the predictive model. A discriminant analysis will classify the level of caries severity and the performance of the model will be further discussed.

II. METHODOLOGY

Flowchart of procedures involved in this study were illustrated in Fig. 1.

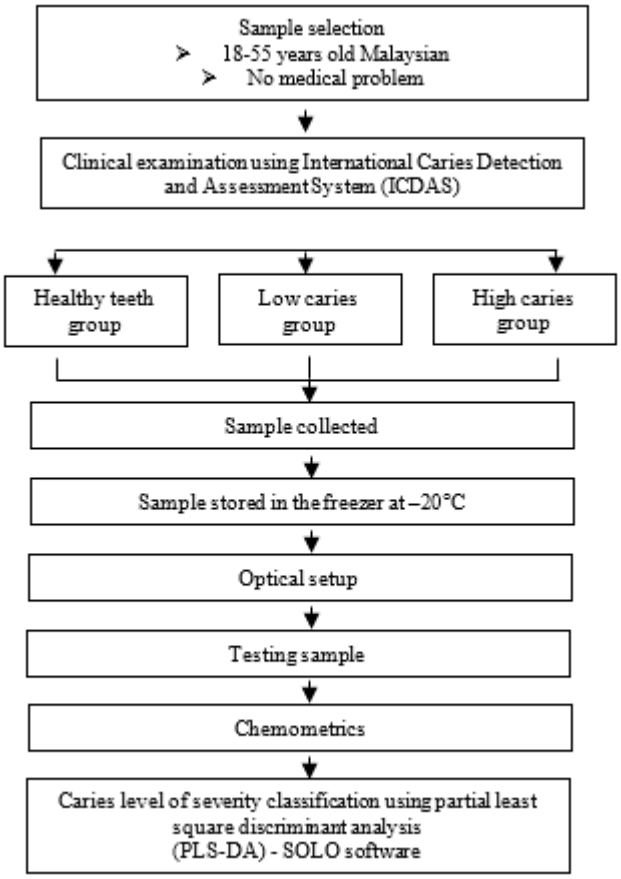

Fig. 1. Flowchart diagram of the research methodology.

A. Subject selection

The samples were acquired from the Dental Clinic, Faculty of Dentistry, Universiti Kebangsaan Malaysia with ethics code UKM PPI/111/8/JEP-2018-441. The inclusion criteria of patients selected from this study are Malaysian aged 18-55 years old with no underlying medical problems. The International Caries Detection and Assessment System (ICDAS) was used to conduct a clinical examination for the detection of caries levels. Teeth with ICDAS 0 are categorized as healthy teeth with no presence of caries (sound teeth), the teeth with ICDAS 1, 2 and 3 as low caries group, and teeth with ICDAS 4 and 5 as high caries group.

B. Sample Preparation Process

The Saliva Collection Aid (SCA) kit was used to gather entire saliva using the passive drool method. The saliva is collected using a saliva collecting tube. A protease inhibitor is added to the saliva samples. The saliva sample was maintained at 4°C for 15–25 minutes before being transferred to the centrifuge (Heraeus Fresco 21 centrifuge, Thermo Scientific). To achieve pure saliva and avoid light scattering during testing, the samples were centrifuged for 30 minutes at 6500 rpm and 4°C. The samples are then kept at –20°C at the proteomic laboratory at UKM's Faculty of Dentistry's cryostorage box. The samples are then thawed at room temperature before being characterised.

C. Optical Setup

The deuterium lamp, which is housed in a unique housing, produces ultra violet light with wavelengths ranging from 160 nm to 500 nm. A unique lens collimates the UV spectrum and couples it to a bifurcated optical fibre probe. The probe is a waveguide with a 1000 m fibre core diameter and a 6-around-1 fibre bundle configuration. Six core fibres supply UV light to the sample, with one fibre in the middle collecting reflected light and returning it to the other end. The sample is placed in a custom cuvette holder and placed in an unique 350 L quartz cuvette that facilitates UV spectrum transmission. A high-resolution UV-VIS spectrometer measures reflected UV light with a wavelength detection range of 200 nm to 1100 nm at 0.01nm resolution.

D. Chemometrics

The UV spectra collected will be trained using partial least square discriminant analysis (PLS-DA) to classify the caries level of severity using SOLO software (Eigenvector). A total of 102 data obtained is split into calibration and validation with ratio of 80:20 using Kennard Stone algorithm. The number of variable are 561. Preprocessing such as autoscale and multiplicative scatter correction (MSC) will be applied on the spectra. The latent variable, LV set for the PLS-DA was 5.

III. RESULT AND DISCUSSION

The saliva sample was irradiated with UV radiation and the concentration of saliva content will absorb certain amount energy that follow the Beer Lambert law (refer equation 1). The UV spectra obtained as illustrated in Fig. 2.

$$A = \varepsilon c l \tag{1}$$

Where A is the absorbance, ε is the analyte molar absobtivity, c is the concentration of the compound and l is the path length of the light travelled [16].

Peak absorbance measurements around 280 nm were detected by the spectrometer. According to earlier research, peak absorption is associated with the presence of aromatic amino acids such as tyrosine and tryptophan in salivary alpha-amylase, which serves as a binding site for oral bacteria. [14]. In proteins, the indole group of tryptophan is thought to be the main source of ultraviolet absorbance at 280 nm [15]. The spectra undergone chemometrics process to extract the information related to the alpha-amylase absorption. Based on the spectra shown in Fig. 2, the spectra was embedded with noise thus preprocess are required to enhance the signal-to-noise ratio. Autoscale method will mean centre the data and divide it by the standard deviation. MSC will correct the data based on the average spectrum. The UV spectra versus the number of variables after the preprocessing as shown in Fig 3.

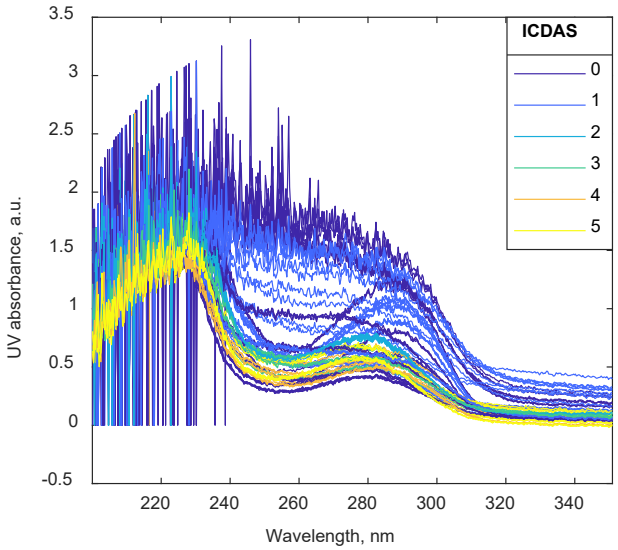

Fig. 2. The UV spectra for saliva samples of different caries severity from ICDAS 0-5.

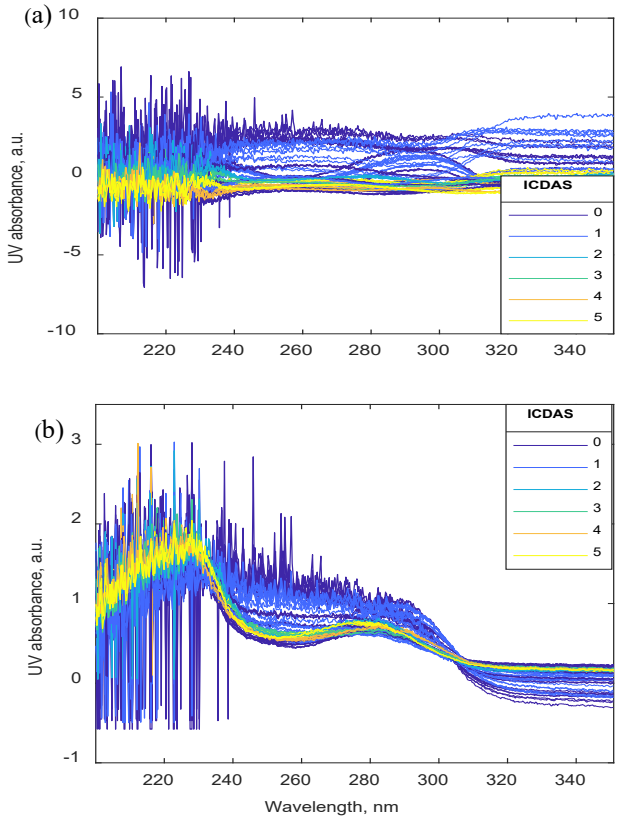

Fig. 3. The UV spectra after preprocessing application, (a) autoscale and (b) MSC.

PLS-DA formed a multivariate relationship between the measurement data matrix X and the ICDAS matrix Y. PLS collects orthogonal features from the spectra and ICDAS matrices to create a new matrices. The new projection of the latent structure based on the latent variables set. The result of the PLS-DA on the spectre without and with preprocessing as tabulated in Table I. The performance of the PLS-DA was evaluated by calculating the sensitivity and specifity.

Sensitivity is a measure of a test's ability to detect true positives, whereas specificity is a measure of its ability to detect real negatives.The true positive, TP is when the model can correctly classifies the sound teeth (ICDAS level 0), the true negative, TN is the ability of the model to predict the caries teeth (ICDAS level 1-5) correctly. The false negative, FN occur when the sound teeth detect as caries whereas the false positive,FP is when the caries teeth predicted as normal teeth. The values of the TP, FP,TN and FN based on the confusion matrix of the measured caries level and the predicted caries level as tabulated in Table I. The calculation of the sensitivity and specificity as stated in equation 2 and 3.

TABLE I. THE VALUES OF TRUE POSITIVE, FALSE POSITIVE, TRUE NEGATIVE AND FALSE NEGATIVE BASED ON THE CONFUSION MATRIX OF THE PLS-DA MODEL FOR DIFFERENT PREPROCESSING.

	TP	FP	TN	FN
None	0.9091	0.0000	1.0000	0.0909
Autoscale	0.9546	0.0000	1.0000	0.0455
MSC	0.9546	0.0000	1.0000	0.0455

	TP	FP	TN	FN
None	0.8000	0.0000	1.0000	0.2000
Autoscale	1.0000	0.1250	0.8750	0.0000
MSC	1.0000	0.0000	1.0000	0.0000

$$Sensitivity = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Negative\ (FN)} \quad (2)$$

$$Specificity = \frac{True\ Negative\ (TN)}{True\ Negative\ (TN)+False\ Positive\ (FP)} \quad (3)$$

TABLE II. THE RESULT OF PLS-DA IN TERMS OF THE SENSITIVITY AND THE SPECIFICITY FOR NONE, AUTOSCALE AND MSC PREPROCESSING.

Parameter	None	Autoscale	MSC
Sensitivity (Calibration)	0.96	0.96	1.00
Sensitivity (Validation)	0.80	1.00	1.00
Specificity (Calibration)	1.00	1.00	1.00
Specificity (Validation)	1.00	1.00	1.00

The result of the sensitivity and specificity for no preprocessing were reported for calibration and validation. The sensisitivity for none preprocessing during calibration was 0.96 whereas for validation was 0.80. The specificity for none preprocessing for calibration and validation were 1.00 respectively. After the autoscale preprocessing, the sensitivity value for the validation has improved from 0.80 to 1.00 but the sensitivity for the calibration remain unchanged. The implementation of the MSC has shown the best result because

the values of the sensitivity and specificity for calibration and validation were 1.00. This result proves that the data preprocessing has optimized the performance of the PLS-DA model.

The performance of the MSC with PLS-DA can be measure with the receiver operating charateristics (ROC) curve. Fig. 4 shows the plot of sensitivity versus 1-specificity. The area under the curve (AUC) was computed to evaluate the overall classification performance of the PLS-DA. The AUC value obtained was 1.00. The result shows that PLS-DA with MSC preprocessing can work as the predictive model to detect the presence of dental caries.

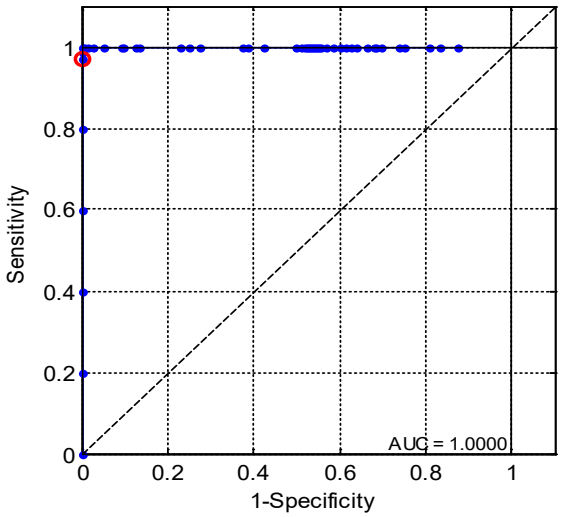

Fig.4. The ROC of the PLS-DA model after the MSC preprocessing.

IV. Conclusion

The UV absorption spectroscopy combined with chemometrics has shown a good result in classifying the level of caries severity using saliva sample. The PLS-DA model combined with few data preprocessing has shown optimized result of the classification model. The best value of the sensitivity and the specificity obtained were 1.00 with MSC preprocessing. The ROC plot gives 1.00 value for the AUC thus proves the capability of the model to classify the caries level of severity using ICDAS.

Acknowledgment

The authors would like to acknowledge with gratitude the Photonics Department Laboratory, MIMOS and Faculty of Dentistry, Universiti Kebangsaan Malaysia (UKM) for the facilities. This research was financially supported by the UKM Geran Universiti Penyelidikan (GUP-2017-088).

References

[1] J. C. Carvalho and U. Schiffner, “Dental Caries in European Adults and Senior Citizens 1996-2016: ORCA Saturday Afternoon Symposium in Greifswald, Germany - Part II,” *Caries Res.*, vol. 53, no. 3, pp. 242–252, 2019, doi: 10.1159/000492676.

[2] D. T. Lui, A. M. R.Wahab, E. Kuppusamy, N. H. Hamzaid, M. R. Hassan, and F. Yazid, "Association of Early Childhood Caries and Nutritional Status: A Scoping Review," *J Clin Pediatr Dent*, vol. 47, no. 3, 2023.

[3] A. Mira, A. Artacho, A. Camelo-castillo, S. Garcia-esteban, and A. Simon-soro, “Salivary Immune and Metabolic Marker Analysis (SIMMA): A Diagnostic Test to Predict Caries Risk,” *Diagnostics*, vol. 7, no. July, 2017, doi: 10.3390/diagnostics7030038.

[4] P. Ahmad, A. Hussain, A. Carrasco-Labra, and W. L. Siqueira, "Salivary Proteins as Dental Caries Biomarkers: A Systematic Review," *Caries Res.*, vol. 56, no. 4, pp. 385-398, 2022.

[5] K. N. Basri, F. Yazid, R. M. A.Wahab, M. N. M. Zain, Z. M. Yusof, and A. S. Zoolfakar, "Chemometrics Analysis for the Detection of Dental Caries via UV Absorption Spectroscopy," *Spectrochim. Acta A Mol. Biomol. Spectrosc.*, vol. 266, p.120464, 2022.

[6] J. Choi, H. Eun, and C. Kim, “Boosting Proximal Dental Caries Detection via Combination of Variational Methods and Convolutional Neural Network,” *J. Signal Process. Syst.*, vol. 90, no. 1, pp. 87–97, 2018, doi: 10.1007/s11265-016-1214-6.

[7] S. H. X. Tan, Y. Y. Teo, M. H. X. Tan, and X. Gao, “Childhood Factors and Dental Caries in the Permanent Dentition: Findings of an 8-Year Study Under a Nationwide School Dental Service,” *Int. Dent. J.*, vol. 0, pp. 1–8, 2021, doi: 10.1016/j.identj.2021.01.008.

[8] M. B. Diniz, J. A. Rodrigues, K. W. Neuhaus, R. C. L. Cordeiro, and A. Lussi, “Influence of examiner’s clinical experience on the reproducibility and accuracy of radiographic examination in detecting occlusal caries,” *Clin. Oral Investig.*, vol. 14, no. 5, pp. 515–523, 2010, doi: 10.1007/s00784-009-0323-z.

[9] C. Guang and B. Lin, “Quantitative method to assess caries via fluorescence imaging from the perspective of autofluorescence spectral analysis,” no. October, 2015, doi: 10.1088/1054-660X/7/8/085601.

[10] N. Miyamoto *et al.*, “Molecular Fingerprint Imaging to Identify Dental Caries Using Raman Spectroscopy,” *Materials (Basel).*, 2020.

[11] P. Usenik, M. Bu, and F. Pernus, “Near-infrared hyperspectral imaging of water evaporation dynamics for early detection of incipient caries,” *J. Dent. 42*, vol. 2, pp. 1242 – 1247, 2014.

[12] N. Abogazalah, G. J. Eckert, and M. Ando, “In vitro performance of near infrared light transillumination at 780-nm and digital radiography for detection of non-cavitated approximal caries,” *J. Dent.*, vol. 63, no. May, pp. 44–50, 2017.

[13] K. Cheng, P. She, H. Wang, Z. Wang, L. Zhang, X. Tang, L. Yuan, Y. Feng, X. Song, G. Pan, and J. Yang, "A Bio-Inspired Versatile Free-Standing Membrane for Oral Cavity Microenvironmental Monitoring and Remineralization to Prevent Dental Caries," *Mater. Horiz.*, vol. 10, no. 2, pp. 512-523, 2023.

[14] M. A. Adebayo, A. N. Kolawole, B. A. Falese, and A. O. Kolawole,. "Spectroscopic and In Silico Evaluation of Hesperetin, Aglycone Flavanone, as a Prospective Regulatory Ligand for Human Salivary α-Amylase," *J. Biomol. Struct. Dyn.*, pp. 1-16, 2023.

[15] L. Huang, X. Zhang, and Z. Zhang, “UV – vis sensor array combining with chemometric methods for quantitative analysis of binary dipeptide mixture (Gly-Gly / Ala-Gln),” *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 221, 2019.

[16] T. Liang and R. Roy, “Ultraviolet-Visible Spectrophotometry (UV-VIS) and SALIgAE® Qualitative and Semi-Quantitative Tools for the Analysis of Salivary Amylase,” *J. Forensic Res.*, vol. 5, no. 5, 2014, doi: 10.4172/2157-7145.1000247.