IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 January 2023, accepted 15 February 2023, date of publication 16 February 2023, date of current version 23 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3246299

RESEARCH ARTICLE

Absorption Spectroscopy in Dental Tissue Analysis

ALEŠ PROCHÁZKA[1,2], (Life Senior Member, IEEE), DANIEL MARTYNEK[1], AND JINDŘICH CHARVÁT[3]

[1]Department of Mathematics, Informatics and Cybernetics, University of Chemistry and Technology, Prague (UCT Prague), 160 00 Prague, Czech Republic
[2]Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague, 160 00 Prague, Czech Republic
[3]Department of Stomatology, First Faculty of Medicine, Charles University in Prague, 116 36 Prague, Czech Republic

Corresponding author: Aleš Prochazka (A.Prochazka@ieee.org)

This work was supported by Charles University under Project GA UK 52220.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of the General University Hospital, Prague, under Approval No. 613/18 S-IV, and performed in line with the Helsinki Declaration.

ABSTRACT Oral health problems are closely associated with the analysis of dental tissue changes and the stomatologic treatment that follows. This paper explores the use of diffuse reflectance spectroscopy in the detection of dental tissue disorders. The data set includes 343 measurements of teeth spectra in the wavelength range from 400 to 1700 nm. The proposed methodology focuses on computational and statistical methods and the use of these methods for the classification of dental tissue into two classes (healthy and unhealthy) by estimating the probability of class membership. Signal processing is based on the difference between the healthy and unhealthy teeth reflectance spectra in the infrared and visible ranges. Selected features associated with observed spectra are then used for machine learning classification based on the experience of an expert in stomatology during the learning stage. The proposed modification of the weighted *k*-nearest neighbour method provides class boundaries and the probability of class membership during the verification stage. The accuracy of the classification process reached 95.4%. The proposed methodology and graphical user interface point to the possibility of using absorption spectroscopy in the evaluation of tissue quality changes and its possible implementation in the clinical environment.

INDEX TERMS Diffuse reflectance spectroscopy, dental tissue changes, feature extraction, machine learning, weighted *k*-nearest neighbour method, computational intelligence, classification.

I. INTRODUCTION

Dental caries (also known as tooth decay) is the most common oral disease; it results in the demineralisation of the enamel and dentin as it progresses. It is estimated that 3.5 billion people suffer from oral diseases; the majority of them suffer from dental caries. The associated problems can cause tooth loss and pain, and they may induce a series of oral and general diseases. The early detection of the demineralisation process is thus important for treatment planning.

Caries detection methods usually depend on the subjective assessment of the examining physician. To prevent problems with dental caries [1], [2], [3], [4], dental lesions should be identified as soon as possible so that minimally invasive treatment procedures can be implemented [5]. Besides the most common visual-tactile method, further procedures, including diffuse reflectance spectroscopy, quantitative light-induced fluorescence, and vibrational spectroscopic techniques [6], [7], [8], [9], [10], can be applied for studies of the structural or elemental features of hard dental tissues. These methods offer safer, objective, and reliable options that can be applied to patients without X-ray exposure.

Enamel (the outermost tissue of teeth) is a protective layer. It is the hardest substance in the human body, and it consists mostly (up to 95–96%) of inorganic salts (hydroxyapatite). The main part (dentine) of the tooth contains water and organic salts [11], [12], [13]. The cause of dental caries is related to a decrease in the pH and the chemical

The associate editor coordinating the review of this manuscript and approving it for publication was Ghulam Muhammad.

decomposition of hydroxyapatite (HPA); this causes the loss of a protective layer and then the damage of the organic structure. The decomposition of HPA is initiated by low pH values (caused by bacteria metabolism, which creates weak organic acids) or low concentrations of calcium or phosphate in saliva [11]. A healthy tooth will have a different content than an unhealthy tooth, i.e. an unhealthy tooth will have less HPA and more water. Therefore, we can assume that different spectra will be obtained from healthy and unhealthy teeth [14]. More precisely, at wavelengths of around 600 nm, we can observe more reflected light for healthy teeth because healthy enamel (with the highest HPA content) has a reflectance maximum around this wavelength [15].

Diffuse reflectance spectroscopy (DRS) [16], [17], also known as reemission spectroscopy, is a non-invasive method that measures the reflectance spectrum of light passing through and reflecting off of a material. Two main factors play a role – the absorbance and scattering of light [18]. This method can be used to determine whether a tooth has caries or not. As an unhealthy tooth has fewer minerals, small dints, and some pigmentation, the reflectance spectrum of these areas can be different from that of a healthy tooth. Therefore, it may be possible to determine whether tooth tissue is sound or not using this method.

Mathematical methods used for dental data processing [6], [19], [20] include general methods for signal and image analysis [21], [22], digital filtering [23], segmentation, and feature estimation. Machine learning methods and deep learning tools [24], [25], [26], [27] can then be applied for the classification [28] of individual samples. Convolutional neural networks and computational methods for automated dental tissue analysis [29], [30], [31] are often used in this area.

Some previous studies [32] describe the use of optical properties of affected tissues and spectroscopic detection of caries lesions, but the wavelength range is narrower in some cases and accuracy of diagnostics is lower as well. Further studies [33] are devoted to the use of spectroscopic methods and early diagnostics for avoiding more invasive treatment and they point to the necessity of a full-scale imaging system to identify caries lesions that are not detected by traditional visual examination. Spectroscopic techniques [34], [35] belong to emerging technologies for dentin caries detection and surface analysis of dental materials. Present limitations are in the necessity to work with properly prepared samples without water that strongly absorbs radiation [36].

The present paper forms a contribution to this research using a wide wavelength range, selection of absorption spectroscopy data features, optimization of their number, evaluation of probabilities of the class membership, and proposal of new computational methods for spectroscopic data and their analysis. It is devoted to the analysis of 343 samples; each of them is analysed ten times to find each tooth's spectrum in the visible and infrared ranges. The proposed methodology is based upon the classification of selected data features and the use of a machine learning method to find a computational system for the estimation of the probability of class membership for a new sample to detect tissue changes in their early stage.

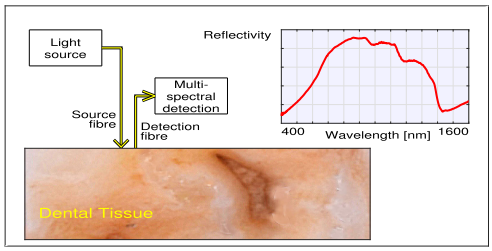

FIGURE 1. Principles of the diffuse reflectance spectroscopy method applied to the measurement of the dental tissue reflectivity associated with different wavelengths of the source light and used for tissue quality estimation.

II. METHODS

A. DATA ACQUISITION

All data were acquired and manually classified by an experienced stomatologist, the last author of the present paper. The whole dataset includes 343 samples, with 10 records of reflectance spectroscopy acquired for each sample. Detailed descriptions of each observation, including visual classification into two classes (healthy and unhealthy tissues), are contained in the associated spreadsheet along with other information used during the computational processing stage. These procedures involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments.

Figure 1 presents the principles of diffuse reflectance spectroscopy. The light of defined spectral components is carried from a light source through the optical fibre to the tooth tissue, where the light is absorbed, scattered, and reflected. Another optical fibre carries the reflected light to a spectrometer (or many spectrometers via a fibre splitter). A spectrometer evaluates a spectrograph that points to features of the selected body surface area.

A tungsten-halogen broadband light source with an integrated shutter (Avantes HAL-S) was used as the light source. The light was transmitted with a 1.6-mm-diameter fibre-optic probe, with the distal end polished at an angle of 5 degrees. The probe consisted of two 200-μm-core-diameter optical fibres with a tip-to-tip distance of 1.2 mm. The illumination fibre was connected to the light source and the collection fibre was connected to a visible light spectrometer with a silicon detector (OceanOptics Maya 2000PRO) and a near-infrared spectrometer with an InGaAs detector (Horiba-S330-2 NIR) via a 50–50 % fibre splitter that was used to divide the collected light between the two detectors. The acquisition time per spectrum was 1 s including background measurement.

B. SIGNAL PREPROCESSING

The resulting signal can be further filtered and normalized for the processing stage. Its features (wavelength values at the reflectance maximum, minimum reflectance, etc.) are then extracted.

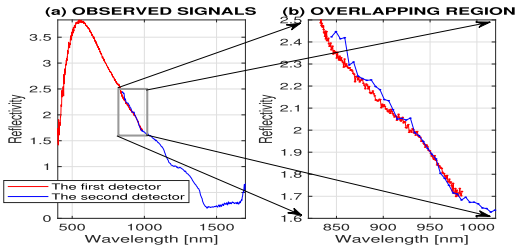

FIGURE 2. Selected reflectance signal recorded by two spectrometers covering different ranges and wavelength resolutions: (a) observed signals and (b) detailed values in the overlapping area.

The mathematical methods of the data processing procedure are closely related to the properties of the sensors used for the acquisition of the data. However, in general, signal de-noising and the extraction of features in both the time and frequency domains create common problems.

Each signal $\{x(n)\}_{n=0}^{N-1}$ of spectral values was analyzed at first to eliminate gross measurement errors and then smoothed by the low-pass finite impulse response (FIR) filter of order $M = 30$ defined by the following relation:

$$y(n) = \sum_{k=0}^{M-1} b(k)\, x(n-k). \tag{1}$$

A selected normalised cutoff frequency of 0.005 was used to evaluate a new sequence $\{y(n)\}$ for all values of $n = 0, 1, 2, \cdots, N-1$ and for filter coefficients $\{b(k)\}_{k=0}^{M-1}$.

C. FEATURE DESCRIPTION

The observations of two detectors recorded with non-uniform sampling overlap in the range of wavelengths $\langle 844, 984\rangle$ nm, as shown in Fig. 2 for a selected measurement. The mean wavelength resolutions of the first detector (up to 984 nm) and the second detector (from 844 nm) were 1.19 nm and 3.49 nm, respectively. Polynomial fitting, resampling, and weighted averaging are applied in this range to eliminate duplicate reflectance values and to enable the further processing of the reflectance observations.

The following classification process was based on the signal features that differentiate healthy and unhealthy spectra. The following features were selected:

1) The wavelength of the maximum reflectance (which is lower for healthy teeth);
2) The wavelength of the minimum reflectance;
3) The upward slope at low wavelengths (which is higher for healthy teeth);
4) The downward slope after the maximum reflectance (which is lower for the reflectance signals of healthy teeth).

These features formed a column feature vector p_j for each observation $j = 1, 2, \cdots, Q$.

A common problem of pattern recognition and machine learning is in the selection of the most relevant features to reduce the learning process time and to enable visualization of data clustering. To find the optimal pair of features $f1$ and $f2$, a specific criterion was constructed. It is based on the Euclidian distance between the centres of gravity of the features of healthy $(A_{f1,f2})$ and unhealthy $(B_{f1,f2})$ tissues and their standard deviations, and it is meant to detect well-separated and compact clusters. The numerical comparison of clusters was performed using the following criterion:

$$C(f1, f2) = \frac{dist(M(A_{f1,f2}), M(B_{f1,f2}))}{mean(S(A_{f1,f2})) + mean(S(B_{f1,f2}))}, \tag{2}$$

where $M(c)$ and $S(c)$ represent mean values and standard deviations of clusters associated with classes c_A and c_B, respectively. During the processing stage, the highest value of this criterion was found for features $f1$ and $f2$, which represent the wavelength value at maximum reflectance and the initial gradient value of the observed spectrum.

Standard methods of identifying the most important features are often based on singular value decomposition [37], QR factorization with Gram-Schmidt orthogonalization process, and principal component analysis [38].

D. CLASSIFICATION

The k-nearest neighbour (kNN) method was applied, as it is the machine learning algorithm that is most frequently used to classify the features of Q pattern vectors in which there is little or no information about the data distribution [39]. This method finds k nearest neighbours for each (reference) data point r, specified by the column vector $\{r(i, 1)\}_{i=1}^{R}$ of its R features, that should be classified. This method needs a set of column feature vectors $(p_1, p_2, \ldots, p_j, \ldots p_Q)$, and each of them must be associated with a target class $(T_1, T_2, \ldots, T_j, \ldots, T_Q)$ defined by an expert in order to form the training set. Each feature vector p_j includes R features $\{p(i, j)\}_{i=1}^{R}$ that form the feature matrix $P_{R,Q}$. This method computes the Euclidean distance (in most cases) between the feature vector that will be classified and every feature vector from the training set. Then, k vectors from the training set that have the lowest Euclidian distances to the vector are selected, their class memberships are found, and the majority of these classes are assigned to the reference data point class.

A proposed modification of the weighted k-nearest neighbours method [40], [41], [42] was used in the present paper for the classification of spectral features, with the main difference in the last (classification) step. The weighted kNN method utilises the absolute values of the distances $\{d_r(i)\}_{i=1}^{k}$ between the locations of the selected feature vector r and its k nearest neighbours. In the present study, the weighting function is

$$w_r(i) = \frac{1}{d_r(i)}, \tag{3}$$

where $w_r(i)$ is the vector of weight values associated with k of the vector's neighbours for $i = 1, 2, \ldots, k$. The probability that a selected tissue specified by the feature vector $\{r(i, 1)\}_{i=1}^{R}$ is associated with class c is then estimated using the following relation:

$$R_r(c) = \frac{\sum_{i \epsilon \phi(c)} w_r(i)}{\sum_{i=1}^{k} w_r(i)}, \tag{4}$$

where $\phi(c)$ is the set of indices of the weighting function $w_r(i)$ associated with the neighbours belonging to class c.

The k-fold cross-validation method was used to determine the ability of a predictive model to perform the classification during its practical implementation. This algorithm is based upon the partitioning of the data set values into k subsets (folds) [43]. Then, the algorithm treats each fold as a test set and the rest of the folds as a training set. Each feature vector from the test set is classified by using data from the training set to determine how accurate the classification is. The cross-validation error is then evaluated as a fraction of the incorrectly determined target classes and the number of pattern values.

The classification algorithm performs the detection of the following two classes in this case: healthy (negative) and unhealthy (positive) tissues. In this case, the results can be distributed into four categories – true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The cross-validation error is then expressed by the following relation:

$$CVE = \frac{FP + FN}{FP + FN + TP + TN}. \tag{5}$$

In the present paper, the leave-one-out method is used as a special case of k-fold cross-validation. The peculiarity of this method is that the number of folds k is the same as the number of data points in one data set. This means that the leave-one-out method takes each data point from the data set and tries to classify it using the rest of the data set. The accuracy of the algorithm is calculated as the ratio of well-classified data points to all data points:

$$ACC = \frac{TP + TN}{FP + FN + TP + TN}. \tag{6}$$

The sensitivity (*true positive ratio*) and specificity (*true negative ratio*) are defined by the following relations:

$$SENS = \frac{TP}{TP + FN}, \qquad SPEC = \frac{TN}{TN + FP}. \tag{7}$$

They quantify the ability of the method to detect all positive or negative values, respectively [44].

During the learning stage, the parameters of the selected classification model are optimised. Then, the final model can be used to process the newly observed data points to predict the classes that they belong to.

III. RESULTS

The database of spectral measurements includes observations of 343 dental tissues; each of these tissues was measured ten times.

Figure 3 presents the resulting reflectance signals in the range of $\langle 400, 1700 \rangle$ nm for both healthy and unhealthy tissues, with their mean values evaluated for the individual classes of healthy and unhealthy tissues. All signals acquired during the experiments were preprocessed, classified, and analysed in the computational environment of MATLAB 2022a.

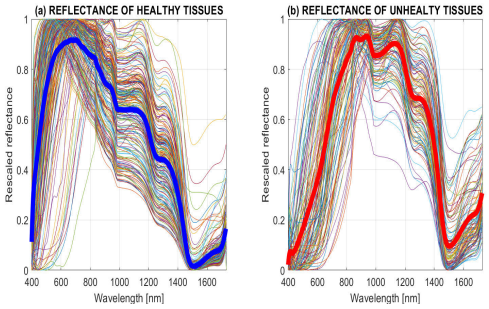

FIGURE 3. Observed signals showing (a) the spectra of healthy teeth, (b) the spectra of teeth with caries, and the average reflectance values of both classes.

TABLE 1. Criterion values of observed reflectance signals for selected features (1: the wavelength of the maximum reflectance, 2: the wavelength of the minimum reflectance, 3: the upward slope at low wavelengths, 4: the downward slope after the maximum reflectance).

Feature f1	*Feature f2*			
	1	2	3	4
1		1.37	**1.92**	0.61
2	1.37		1.36	0.66
3	**1.92**	1.36		0.75
4	0.61	0.66	0.75	

The signal preprocessing procedure included the following steps:

- The evaluation of the averages of measurements associated with one experiment;
- The estimation of reflectance values using the information about the raw intensity, background intensity, and needle calibration;
- The selection of observations in the range of $\langle 400, 1700 \rangle$ nm, as observations outside this range of values recorded by the spectrometers were not correct; resampling to 921 wavelengths followed to enable further processing;
- The construction of the data matrix $D_{R,Q}$, which included R reflectance values for each of the $Q = 343$ observations, and the target vector $T_{1,Q}$ containing the associated classes, which were specified by an experienced stomatologist;
- The elimination of records with gross measurement errors using the variance of the squared difference between average and individual spectra in each class; using a selected threshold value, 11 healthy teeth spectra and 5 spectra from teeth with caries were removed. This removal reduced the size of the data set by 4.7% to 327 spectra that include 211 healthy and 116 unhealthy tissues.
- The detection and elimination of isolated defects (spikes) inside the observed signal and the digital filtering of separate signals using a low-pass FIR filter;

The feature matrix $P_{4,Q}$ included one of these features in each row for all samples. The following analysis using criterion 2 provided the results shown in Table 1. This criterion indicated that feature 1 (the wavelength at the maximum reflectance) and feature 3 (the upward slope) had the most compact and best-separated cluster centres.

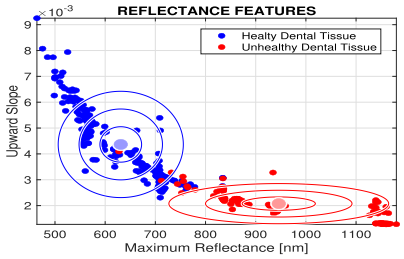

FIGURE 4. The distribution of two features (the wavelength at the reflectance maximum and the upward slope value) with the centre of mass of each class (healthy and unhealthy teeth) and with limits showing 0.5, 1, and 1.5 times the standard deviation.

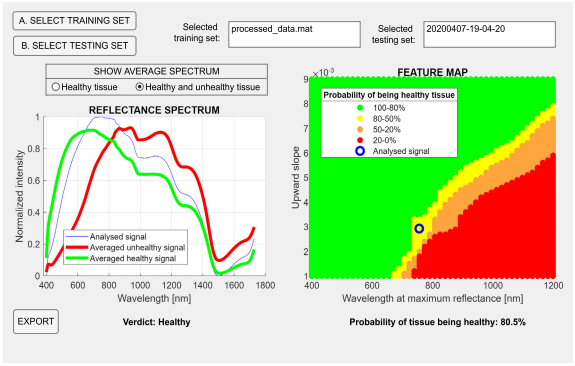

FIGURE 5. Graphical user interface used for the classification of stomatologic tissues, the locations of the selected tissue features, and the evaluation of their class membership probabilities.

The selection of the two most relevant features was confirmed by the singular value decomposition of matrix **A** formed as a transposition of matrix **P** and by evaluation of its singular values ordered from the highest to the lowest one. Its first two values pointed to the dominance of two features (63 % and 29 %). The permutation matrix associated with the QR decomposition of matrix **A** was then used to select the most relevant features: the wavelength of maximum reflectance and the initial upward slope.

Figure 4 presents the distribution of two selected features and indicates which of the two classes (healthy teeth and teeth with caries) each data point belongs to; it also includes the centre of gravity of each class and limits showing 0.5, 1, and 1.5 times the standard deviations of patterns associated with individual classes.

Figure 5 presents the proposed graphical user interface (GUI) used for the classification of stomatologic tissues. The whole algorithm is based upon data preprocessed in the initial stage. Tissue features of the selected observations are then evaluated, their locations are presented in the GUI, and the probability of class membership is then estimated. The probability of the class membership is evaluated by the kNN method for each position on the regular grid of the feature map and plotted using the selected colormap. The whole environment is proposed to support the classification of observed data in the clinical environment.

Table 2 presents a comparison of the classification results [21] of a support vector machine (SVM), 3-nearest neighbour method (3NN), Bayesian method [45], and

TABLE 2. Results of classification performed by the support vector machine (SVM), 3-nearest neighbour (3NN), Bayesian, and two-layer neural network methods; the accuracy and cross-validation errors were calculated using the leave-one-out method.

Method	*Accuracy*	*Cross-Validation Error*
SVM Method	0.957	0.046
3NN Method	0.966	0.064
Bayesian Method	0.951	0.052
Neural Network	0.954	0.046

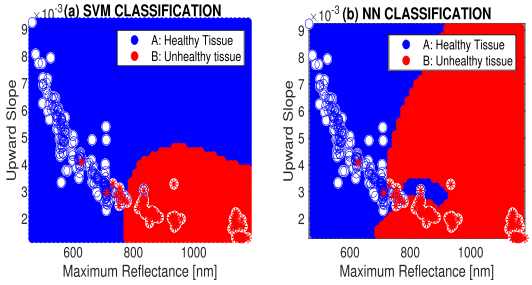

FIGURE 6. Classification of healthy and unhealthy tissues by (a) the support vector machine and (b) the two-layer neural network.

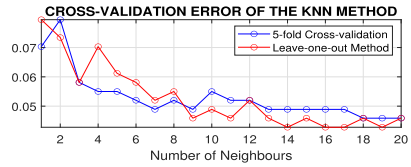

FIGURE 7. Cross-validation error for different number k of neighbours evaluated by the kNN algorithm using 5-fold and leave-one-out method.

two-layer neural network with the sigmoidal and softmax transfer functions using 10 neurons in its first layer. All associated algorithms were created in the computational and visualization environment of MATLAB 2022a with the support of its toolboxes. The accuracy and cross-validation errors were calculated by the leave-one-out method.

The results are very similar for all methods and the given dataset. Figure 6 presents the results of the classification of healthy and unhealthy tissues by the support vector machine method and the two-layer neural network.

Selected datasets are stored at the IEEE DataPort (doi: https://dx.doi.org/10.21227/x9f4-8r14) for further investigation. This repository includes data of healthy and unhealthy tissues, the MATLAB graphical user interface, and the graphical video abstract of the paper.

IV. DISCUSSION

The final data set consisted of 327 data points. These data points were classified using the kNN method with $k = 19$. This value was selected after the proposed analysis of the cross-validation error evaluated for different number of neighbours presented in Fig. 7.

Table 3 presents the number of true positives, true negatives, false positives, and false negatives, along with the calculated sensitivity, specificity, accuracy, and cross-validation (CV) errors evaluated by the leave-one-out method for the data set of 327 tooth spectra using four and the most relevant two features classified by the 3NN and 19NN methods,

TABLE 3. Classification results for the k-nearest neighbour (kNN) method (k=3,19) using leave-one-out method and the pattern matrix formed by N=4 and by the most relevant N=2 features, respectively.

	3NN Method		19NN Method	
	$N=4$	$N=2$	$N=4$	$N=2$
True Positives	207	207	206	206
True Negatives	110	109	106	107
False Positives	6	7	10	9
False Negatives	4	4	5	5
Sensitivity	0.981	0.981	0.976	0.976
Specificity	0.948	0.940	0.914	0.914
Accuracy	0.969	0.966	0.958	0.954
CV Error	0.046	0.064	0.044	**0.042***

Note: * The lowest CV error

TABLE 4. The summary of selected references published in 2020-2022 devoted to detection of dental caries.

Reference	*Labels*	*Proposed Methodology*
Gupta et al. [2]	Raman imaging microspectroscopy	K-mean clustering
Bounds et al. [5]	Near infra-red imaging; absorption; scattering	Image processing; object detection
Shimada et al. [8]	Optical coherence tomography	3D imaging
Orsini et al. [9]	Raman microspectroscopy; vibration spectroscopy	Fourier transform infrared spectroscopy
Cai et al. [10]	Terahertz spectroscopy	Birefringence effect
Charvat et al. [19]	Reflectance spectroscopy	Computational intelligence; classification
Prochazka et al. [24]	Reflectivity data processing	Deep learning; clustering
You et al. [27]	Fluorescence spectroscopy; digital imaging	Deep learning; neural networks
Serban et al. [35]	Optical coherence; spectroscopy; photo-thermal radiometry	Computational intelligence
Kaczmarek et al. [36]	Spectroscopic techniques; Raman spectroscopy	Classification
Manjanga et al. [46]	Radiographic imaging	Segmentation; deep learning; classification
This paper	Absorption spectroscopy	Feature selection classification

respectively. Results suggest that the selection of two features is efficient enough.

Absorption spectroscopy belongs to optical methods used for the detection of dental caries in the oral cavity. These methods are based on light scattering, absorption, and fluorescence as summarized in Table 4 referring to latest references. The associated optical spectrum includes information about biochemical structure of a tissue. Spectral imaging [46] forms a promising approach for diagnostics of tissue diseases and tartar formation. Specific image processing tools, segmentation [6], image components labelling, and deep learning technologies should be included in further studies.

V. CONCLUSION

The paper is devoted to the study of the use of diffuse reflectance spectroscopy in dental tissue analysis. Selected mathematical methods were used to classify healthy dental tissues and dental tissues with caries. The set of 343 teeth spectra was analysed in the visible and low-infrared ranges; the number of teeth spectra in the data set was reduced by 16 due to gross measurement errors.

The features of the spectra were analysed using selected statistical methods. The best features for further classification included the wavelength at maximum reflectance and the upward slope at the beginning of the spectra. The observed signals were used as a training set for the weighted k-nearest neighbours classification method. The statistical analysis of the features resulted in the selection of 19 neighbours.

The graphical user interface for classifying teeth spectra into two classes (healthy and unhealthy) was created and analysed using k-fold cross-validation and the leave-one-out method, respectively.

These results suggest that this methodology could help in identifying teeth with caries. Further studies should be devoted to more complex classification methods, including deep learning and artificial intelligence; the use of these methods could eliminate the need to select the most efficient set of features.

REFERENCES

[1] F. Casalegno, T. Newton, R. Daher, M. Abdelaziz, A. Lodi-Rizzini, F. Schürmann, I. Krejci, and H. Markram, "Caries detection with near-infrared transillumination using deep learning," *J. Dental Res.*, vol. 98, no. 11, pp. 1227–1233, Oct. 2019.

[2] S. D. Gupta, M. Killenberger, T. Tanner, L. Rieppo, S. Saarakkala, J. Heikkilä, V. Anttonen, and M. A. J. Finnilä, "Mineralization of dental tissues and caries lesions detailed with Raman microspectroscopic imaging," *Analyst*, vol. 146, no. 5, pp. 1705–1713, 2021.

[3] J.-H. Lee, D.-H. Kim, S.-N. Jeong, and S.-H. Choi, "Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm," *J. Dent*, vol. 77, pp. 106–111, Oct. 2018.

[4] D. Martynek, "Diffuse reflectance spectroscopy in dental tissue analysis," Bc thesis, Dept. Comput. Control Eng., Univ. Chem. Technol., Prague, Capital of the Czech Republic, Jun. 2022.

[5] A. D. Bounds and J. M. Girkin, "Early stage dental caries detection using near infrared spatial frequency domain imaging," *Sci. Rep.*, vol. 11, no. 1, p. 2433, Jan. 2021.

[6] M. Yadollahi, A. Procházka, M. Kašparová, and O. Vyšata, "The use of combined illumination in segmentation of orthodontic bodies," *Signal, Image Video Process.*, vol. 9, no. 1, pp. 243–250, Jan. 2015.

[7] J. Gomez, "Detection and diagnosis of the early caries lesion," *BMC Oral Health*, vol. 15, no. S1, pp. 1–7, Sep. 2015.

[8] Y. Shimada, M. F. Burrow, K. Araki, Y. Zhou, K. Hosaka, A. Sadr, M. Yoshiyama, T. Miyazaki, Y. Sumi, and J. Tagami, "3D imaging of proximal caries in posterior teeth using optical coherence tomography," *Sci. Rep.*, vol. 10, no. 1, p. 14, Sep. 2020.

[9] G. Orsini, G. Orilisi, V. Notarstefano, R. Monterubbianesi, F. Vitiello, V. Tosco, A. Belloni, A. Putignano, and E. Giorgini, "Vibrational imaging techniques for the characterization of hard dental tissues: From bench-top to chair-side," *Appl. Sci.*, vol. 11, pp. 1–15, Dec. 2021.

[10] J. Cai, M. Guang, J. Zhou, Y. Qu, H. Xu, Y. Sun, H. Xiong, S. Liu, X. Chen, J. Jin, and X. Wu, "Dental caries diagnosis using terahertz spectroscopy and birefringence," *Opt. Exp.*, vol. 30, no. 8, pp. 13134–13147, 2022.

[11] X. Li, J. Wang, A. Joiner, and J. Chang, "The remineralisation of enamel: A review of the literature," *J. Dentistry*, vol. 42, pp. S12–S20, Jun. 2014.

[12] A. B. Mann and M. E. Dickinson, "Nanomechanics, chemistry and structure at the enamel surface," *Monographs Oral Sci.*, vol. 19, pp. 105–131, Jan. 2006.

[13] Y. Elkeraemy, "Application of Raman spectroscopy on teeth," M.S. thesis, Univ. California, Dept. Chem., Riverside, CA, USA, May 2022.

[14] L. Karlsson, "Caries detection methods based on changes in optical properties between healthy and carious tissue," *Int. J. Dentistry*, vol. 2010, Apr. 2010, Art. no. 270729.

[15] A. Kishen, A. Shrestha, and A. Rafique, "Fiber optic backscatter spectroscopic sensor to monitor enamel demineralization and remineralization in vitro," *J. Conservative Dentistry*, vol. 11, pp. 63–70, Apr. 2008.

[16] F. Tetschke, F. L. Kirsten, J. Golde, J. Walther, R. Galli, E. Koch, and A. C. Hannig, "Application of optical and spectroscopic technologies for the characterization of carious lesions in vitro," *Biomedizinische Technik*, vol. 65, pp. 519–527, Jan. 2018.

[17] C. Ng, E. C. Almaz, J. C. Simon, D. Fried, and C. L. Darling, "Near-infrared imaging of demineralization on the occlusal surfaces of teeth without the interference of stains," *J. Biomed. Opt.*, vol. 24, Mar. 2019, Art. no. 036002.

[18] D. C. G. de Veld, M. Skurichina, M. J. H. Witjes, R. P. W. Duin, H. J. C. M. Sterenborg, and J. L. N. Roodenburg, "Autofluorescence and diffuse reflectance spectroscopy for oral oncology," *Lasers Surg. Med.*, vol. 36, no. 5, pp. 356–364, 2005.
[19] J. Charvát, A. Procházka, M. Fričl, O. Vyšata, and L. Himmlová, "Diffuse reflectance spectroscopy in dental caries detection and classification," *Signal, Image Video Process.*, vol. 14, pp. 1–8, Jul. 2020.
[20] L. Gráfová, M. Kašparová, S. Kakawand, A. Procházka, and T. Dostálová, "Study of edge detection task in dental panoramic radiographs," *Dentomaxillofacial Radiol.*, vol. 42, no. 7, Jul. 2013, Art. no. 20120391.
[21] A. Procházka, O. Vyšata, and V. Mařík, "Integrating the role of computational intelligence and digital signal processing in education: Emerging technologies and mathematical tools," *IEEE Signal Process. Mag.*, vol. 38, no. 3, pp. 154–162, May 2021.
[22] B. Langari, S. Vaseghi, A. Prochazka, B. Vaziri, and F. T. Aria, "Edge-guided image gap interpolation using multi-scale transformation," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4394–4405, Sep. 2016.
[23] E. Hostalkova, O. Vysata, and A. Prochazka, "Multi-dimensional biomedical image de-noising using Haar transform," in *Proc. 15th Int. Conf. Digit. Signal Process.*, Cardiff, U.K., Jul. 2007, pp. 175–179.
[24] A. Procházka, J. Charvát, O. Vyšata, and D. Mandic, "Incremental deep learning for reflectivity data recognition in stomatology," *Neural Comput. Appl.*, vol. 34, no. 9, pp. 7081–7089, May 2022.
[25] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[26] H. Chen, K. Zhang, P. Lyu, H. Li, L. Zhang, J. Wu, and C.-H. Lee, "A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films," *Sci. Rep.*, vol. 9, no. 1, pp. 1–11, Mar. 2019.
[27] W. You, A. Hao, S. Li, Y. Wang, and B. Xia, "Deep learning-based dental plaque detection on primary teeth: A comparison with clinical assessments," *BMC Oral Health*, vol. 20, no. 1, pp. 1–7, Dec. 2020.
[28] Z. Li, S.-H. Wang, R.-R. Fan, G. Cao, Y.-D. Zhang, and T. Guo, "Teeth category classification via seven-layer deep convolutional neural network with max pooling and global average pooling," *Int. J. Imag. Syst. Technol.*, vol. 29, no. 4, pp. 577–583, May 2019.
[29] F. Schwendicke, T. Golla, M. Dreher, and J. Krois, "Convolutional neural networks for dental image diagnostics: A scoping review," *J. Dentistry*, vol. 91, Dec. 2019, Art. no. 103226.
[30] J. Yang, Y. Xie, L. Liu, B. Xia, Z. Cao, and C. Guo, "Automated dental image analysis by deep learning on small dataset," in *Proc. 42nd IEEE Int. Conf. Comput. Softw. Appl.*, Jul. 2018, pp. 492–497.
[31] R. F. Mansour, A. Al-Marghilnai, and Z. A. Hagas, "Use of artificial intelligence techniques to determine dental caries: A systematic review," *Int. J. Neural Netw. Adv. Appl.*, vol. 6, pp. 1–11, Jan. 2019.
[32] M. Ruohonen, K. Palo, and J. Alander, "Spectroscopic detection of caries lesions," *J. Med. Eng.*, vol. 2013, pp. 1–9, Jan. 2013.
[33] E. Yakubu, B. Li, Y. Duan, and S. Yang, "Full-scale Raman imaging for dental caries detection," *Biomed. Opt. Exp.*, vol. 9, no. 12, pp. 1–8, 2018.
[34] Y. Shimada, A. Sadr, Y. Sumi, and J. Tagami, "Application of optical coherence tomography (OCT) for diagnosis of caries, cracks, and defects of restorations," *Current Oral Health Rep.*, vol. 2, no. 2, pp. 73–80, Jun. 2015.
[35] C. Serban, D. Lungeanu, S. D. Bota, C. C. Cotca, M. L. Negrutiu, V. F. Duma, C. Sinescu, and E. L. Craciunescu, "Emerging technologies for dentin caries detection—A systematic review and meta-analysis," *J. Clin. Med.*, vol. 11, pp. 1–23, Jan. 2022.
[36] K. Kaczmarek, A. Leniart, B. Lapinska, S. Skrzypek, and M. Lukomska-Szymanska, "Selected spectroscopic techniques for surface analysis of dental materials: A narrative review," *Materials*, vol. 14, no. 10, p. 2624, May 2021.
[37] M. Afshar and H. Usefi, "Dimensionality reduction using singular vectors," *Sci. Rep.*, vol. 11, no. 1, p. 13, Feb. 2021.
[38] A. Sharma, K. K. Paliwal, S. Imoto, and S. Miyano, "Principal component analysis using QR decomposition," *Int. J. Mach. Learn. Cybern.*, vol. 4, pp. 679–683, Dec. 2013.
[39] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.
[40] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Sci. Rep.*, vol. 12, no. 1, p. 11, Apr. 2022.
[41] M. Bicego and M. Loog, "Weighted K-nearest neighbor revisited," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 1643–1648.
[42] G. F. Fan, Y. H. Guo, J. M. Zheng, and W. C. Hong, "Application of the weighted K-nearest neighbor algorithm for short-term load forecasting," *Energies*, vol. 12, pp. 1–9, Mar. 2019.
[43] C. Sammut and G. I. Webb, "Cross-validation," in *Encyclopedia of Database Systems*. Boston, MA, USA: Springer, 2010.
[44] K. M. Ting, "Sensitivity and specificity," in *Encyclopedia of Machine Learning*. Boston, MA, USA: Springer, 2010, pp. 901–902.
[45] A. Procházka, O. Vyšata, M. Vališ, O. Ťupa, M. Schätz, and V. Marřík, "Bayesian classification and analysis of gait disorders using image and depth sensors of Microsoft Kinect," *Digit. Signal Process.*, vol. 47, pp. 169–177, Dec. 2015.
[46] V. Majanga and S. Viriri, "A survey of dental caries segmentation and detection techniques," *Sci. World J.*, vol. 22, pp. 1–19, Apr. 2022.

ALEŠ PROCHÁZKA (Life Senior Member, IEEE) received the Ph.D. degree from Czech Technical University in Prague, in 1983. He was appointed as a Professor in technical cybernetics with Czech Technical University in Prague, in 2000. He is currently the Head of the Digital Signal and Image Processing Research Group, Department of Computing and Control Engineering, University of Chemistry and Technology, Prague (UCT Prague), and the Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague. His research interests include mathematical methods of multidimensional data analysis, segmentation, feature extraction, classification, and modeling in biomedicine and engineering. He is a member of IET and EURASIP. He has served as an Associate Editor for *Signal, Image and Video Processing* (Springer). He is a reviewer of different IEEE TRANSACTIONS and Springer, Elsevier, and MDPI journals.

DANIEL MARTYNEK received the B.Sc. degree in physical and computational chemistry from the University of Chemistry and Technology, Prague (UCT Prague), Czech Republic, in 2022. He is currently a member of the Digital Signal and Image Processing Research Group, Department of Computing and Control Engineering, UCT Prague. His research interests include information engineering, visualization tools, computational methods of multidimensional data analysis, feature extraction, machine learning, and classification. Applications of his research include biomedicine, neurology, and physiological data processing.

JINDŘICH CHARVÁT received the M.D.Dr. degree from the First Faculty of Medicine, Charles University in Prague, Czech Republic, in 2015. He currently works at the First Faculty of Medicine, Charles University, and the General University Hospital in Prague, and at a private dental practice. His research interests include application of diffuse reflectance spectroscopy in dentistry. The applications of his research include the early diagnosis of dental caries and other tooth-related difficulties. He is a member of the Czech and European Prosthetic Society and is dedicated to prosthodontics in dentistry.