Check for updates

Original Article

Institution of MECHANICAL ENGINEERS
Journal of ENGINEERING IN MEDICINE

Proc IMechE Part H:
J Engineering in Medicine
1–12
© IMechE 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/09544119231162682
journals.sagepub.com/home/pih
SAGE

Development and validation of intraoral periapical radiography-based machine learning model for periodontal defect diagnosis

Fatma Karacaoglu[1], Mehmet Eray Kolsuz[2], Nilsun Bagis[1], Cengiz Evli[2] and Kaan Orhan[2,3]

Abstract

Radiographic determination of the bone level is useful in the diagnosis and determination of the severity of the periodontal disease. Various two- and three-dimensional imaging modalities offer choices for imaging pathologic processes that affect the periodontium. In recent years, innovative computer techniques, especially artificial intelligence (AI), have begun to be used in many areas of dentistry and are helping increase treatment and diagnostic performance. This study was aimed at developing a machine-learning (ML) model and assessing the extent to which it was capable of classifying periodontal defects on 2D periapical images. Eighty-seven periapical images were examined as part of this research. The existence or absence of periodontal defects in the aforementioned images were evaluated by a human observer. The evaluations were subsequently repeated using a radiomics platform. A comparison was made of all data acquired through human observation and ML techniques by SVM analysis. According to the study findings the ability of human observers and the ML model to detect periodontal defects was significantly different in comparison to the gold standard. However, ML and human observers performed similarly for the detection of periodontal defects without a significant difference. This study reveals that the prediction of periodontal defects can be achieved by combining particular radiomic features with image variables. The proposed machine leaning model can be utilized for supporting clinical practitioners and eventually substitute evaluations conducted by human observers while enhancing future levels of performance.

Keywords

Artificial intelligence, deep learning, periapical radiography, alveolar bone defect, periodontal defect

Date received: 12 October 2022; accepted: 19 February 2023

Introduction

Some of the chronic diseases that most frequently impact adults in numerous countries are periodontal diseases.[1,2] Detection of periodontal disease at an early stage is very important in terms of preventing tooth loss and thus protecting the general health of the patient.[3] The diagnosis of periodontitis can be made by clinical and radiographic examinations together. In addition to detailed clinical examination, radiographic determination of the bone level is useful in the diagnosis and determination of the severity of the disease. Radiography is a highly beneficial instrument that can be used for assessing periodontal osseous structures. Various two- and three-dimensional imaging modalities offer choices for imaging pathologic processes that affect the periodontium.

In standard radiographs, the diagnostic information provided by periapical and bitewing projections is the most valuable when evaluating periodontal disease. Along with the technological developments, there has been a tendency toward digital techniques instead of conventional imaging systems in dento-maxillofacial

[1]Department of Periodontology, Faculty of Dentistry, Ankara University, Ankara, Turkey
[2]Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Ankara University, Ankara, Turkey
[3]Ankara University Medical Design Application and Research Center (MEDITAM), Ankara, Turkey

Corresponding author:
Fatma Karacaoglu, Department of Periodontology, Faculty of Dentistry, Ankara University, Ankara 06100, Turkey.
Email: fboke@ankara.edu.

radiology, as in the whole medical community. Charged coupling devices (CCD), complementary metal oxide semiconductor (CMOS) sensors, and phosphor plates (PSP) are used in digital imaging.[4]

Digital subtraction radiography (DSR) is a technique employed for detecting minimal alterations in the dentoalveolar structure by eliminating any structural deterioration in the radiograph and providing the qualitative and quantitative detection of these changes.[5–9] As it is highly sensitive and precise, DSR is utilized in multiple fields including for diagnosing and following-up on the loss of alveolar bone observed in periodontal diseases as well as the evaluation of treatment outcomes.[8,9] However, these two-dimensional radiographs provide no indication of the internal morphology or depth of crater-like defects. Additionally, they offer no information on the degree of involvement on the lingual and facial surfaces.

When compared to periapical radiographs, cone beam computed tomography (CBCT) was able to depict the buccal and lingual/palatal surfaces and allowed the morphology of the periodontal defect to be visualized more effectively.[10] Furthermore, when assessing the bone loss due to periodontal disease, CBCT outperformed digital radiographs in detecting furcation defects, three-wall defects, fenestrations, and dehiscence.[11] However, CBCT imaging is a higher-dose procedure relative to periapical and bitewing radiography and is also more expensive. Therefore, it is suggested that CBCT imaging not be utilized for routinely assessing bone loss due to periodontal disease and should only be considered in selected cases where 2D information is insufficient for diagnostic or treatment planning needs.

In recent years, innovative computer techniques, especially artificial intelligence (AI), have begun to be used in many areas of dentistry and are helping increase treatment and diagnostic demands.[12–15] The concept of "artificial intelligence" denotes the ability of machines to perform tasks normally completed by humans.[16] Different techniques are used in AI like machine learning and deep learning algorithm. Machine learning is one of the AI subfields that involves detecting of patterns in extant data by algorithms, which then engage in self-training to generate predictions based on new information.[17,18] On the other hand, deep learning can perform all machine learning applications, offers functionality with increased complexity, facilitates enhanced decision-making, and is capable of processing massive datasets.[19] Using such algorithms offers multiple benefits including cost-efficiency, time-saving, reduced requirement for human knowledge, and less medical mistakes.[12]

In ML, which is one of the AI subdomains, a machine learns how to complete a given task by mathematically analyzing available datasets instead of being explicitly programed with specific instructions.[20–22] Medical imaging has benefitted from developments in machine learning technology to create an image analysis approach driven by data and based on models with reduced bias from heuristic assumptions about how objects appear within images.[23] There are various studies in the literature focused on mathematical models grounded on AI methods that provide support for specific diagnoses.[24–26]

ML classifiers are algorithm classification techniques widely used in machine learning. K-nearest neighbors (KNN) classification algorithm is one of the most used non-parametric classification methods to determine a sample's classification label based on neighboring samples[27]; support vector machine (SVM) is a computer algorithm that learns by pattern to assign labels to objects[28] while LR is similar to that of linear regression which a classification algorithm that works by determining the similarity of the observed sample with the sample in the model. Decision tree (DT) algorithm is similar to a tree, with branching nodes and leaves which has a specific threshold for classification. In the RF method, DTs with slight differences are combined to form a structure like a forest. By doing so, instead of a result of a single tree, the final decision is achieved by the voting of each tree. Finally, XGBoost is essentially a decision tree with implementation of gradient boosting framework, and can be used for regression, classification, and ranking tasks.[28]

"Radiomics" refers to the high-throughput extraction of large numbers of imaging features, thus converting medical images into mineable high-dimensional data; the subsequent quantitative analysis of these data can support decision-making.[23,29,30] It aims to predict patient-specific outcomes based on high-throughput analysis and mining of advanced imaging biomarkers by machine learning algorithms. Radiomics is a developing area aimed at constructing pertinent statistical models from a significant amount of high-dimensional mineable features taken from medical imaging data (potentially in combination with genomic or clinical data) to facilitate the process of diagnosing disease, prognosis, and monitoring treatment. The radiomics workflow essentially consists of imaging, the region of interest (ROI) or volume of interest (VOI) segmentation, extraction of features, followed by analysis; subsequently, using the chosen features, a statistical model is developed on the basis of machine learning algorithms, which require tuning based on the biological or clinical questions and any a priori knowledge.[23,29–31]

The detecting periodontal defects especially in 3D nature is challenging and plain radiographs are not the preferred method for detecting the alveolar bone defects, implementing radiomics features might be a solution to overcome the inherent limitations of visual evaluation with conventional radiographic techniques. Besides, machine learning has been infrequently applied in the field of dental imaging, with studies only focused on radiomic features, particularly for diagnosing periodontal defects.

Therefore, this study was aimed at developing a machine-learning model and assessing the extent to



Figure 1. Figure showing the periodontal defects on dry mandibles.

which it is capable of classifying periodontal defects on 2D periapical images.

Materials and methods

The study material consisted of eight dry human mandibles with premolar and molar teeth. The mandibles were fixed on specially prepared devices using impression material (Heraeus Kulzer, LLC, 300 Heraeus Way South Bend, USA), while 2 mm pink wax and 1 mm cold acrylic were placed on them as soft tissue equivalents. Attention was paid to the absence of buccal and lingual bone defects in the tooth groups included in the study. The creation of artificial periodontal defects on the molar and premolar teeth of dry skulls was performed individually through the use of burs. Preparation of the defects was then conducted by a periodontal consultant through the use of a high-speed device with hand and angle pieces as well as rounded burs with high volume air/water spray at specific regions chosen on a random basis. The defects were noted to be utilized as the gold standard for radiographic evaluation.

The mandibles were scanned before and after periodontal defects. The first scan was made without defects. Twenty-nine regions were imaged without defects. Subsequently, 2 mm deep defects were created on the alveolar bone walls on the lingual or buccal surfaces of the molars and premolars in each jaw using 1 mm diameter diamond burs (DIATECH, Coltène/Whaledent AG, Switzerland). The buccal defects were created in 29 regions and lingual defects were also created in 29 regions, meaning a total of 58 regions were prepared with defects. After creating the 2 mm defects, digital radiographies were taken again. Following this, the depths of the periodontal defects were increased to 4 mm and the same imaging procedures were repeated. In total, 106 buccal and lingual defects images and 29 defect-free images were obtained for further analysis (Figure 1).

Digital intraoral imaging

The mandibles were scanned before and after periodontal defects. The researchers acquired intraoral radiographs utilizing a Sorodex Digora Optime UV (Sorodex Medical Systems, Helsinki, Finland). Periapical radiographs were acquired using a parallel approach utilizing film holder equipment to ensure standardization. A phosphor plate system functioning with an exposure time of 70 kVp (0.16) was used. Two intraoral radiographs of each defect were taken digitally for the purpose of comparing the quality of the diagnosis and to contribute to our machine learning model.

For the radiographic assessment, a separate evaluation of the individual image set was performed by a single periodontologist (FK-Observer) with 15 years of experience in the field. Calibration of the observers was performed with a PowerPoint presentation consisting of sample defects (MEK) on periapical radiography images for diagnosing periodontal defects.

The non-compressed file format (tagged image file format, TIFF) was used for saving each of the digital intraoral images. DICOM-Viewing software specifically designed for this task was used to display the digital intraoral images (MicroDicom, Sofia, Bulgaria). The optimization of the observation environment was achieved using the identical computer monitor when displaying the images. It was ensured that the observer viewed the monitor at a constant distance of 50 cm, and lighting remained dimmed while the examinations were being conducted. The observer was instructed to determine whether defects were present or not present, in addition to the kind of defects if they existed (Figure 2). The observer was not provided any information on the presence of defects or their exposure characteristics. When conducting the diagnosis, defects were ultimately classified as being present, missing, or it was not possible to detect any.

Patients and dataset

Data management. Radiomics is a developing area that involves the conversion of imaging data into a high-dimensional feature that is capable of being mined through the use of a large volume of automatically extracted data-characterization algorithms. Therefore, in this study, a Radiomics platform (Huiying Medical Technology Co., Ltd, China, http://en.huiyihuiying.com/) was utilized for the purpose of managing the imaging and clinical data, and subsequently to conduct

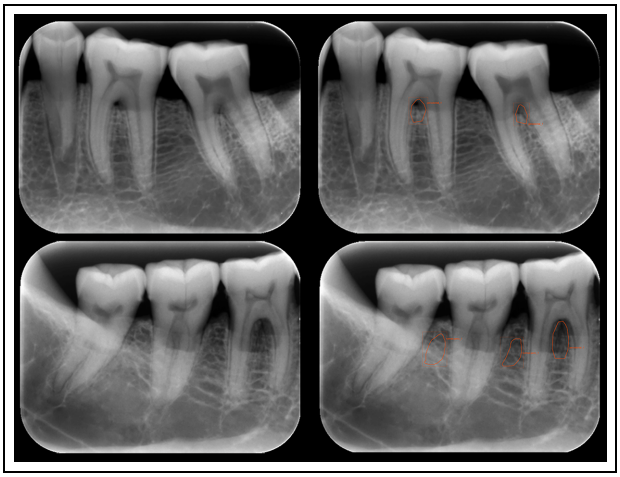


Figure 2. The periapical images with and without periodontal defects together with the selection of VOIs.

radiomics analysis of the statistical features https://www.huiyihuiying.com/en/index.html. Such radiomics platforms can potentially reveal specific imaging algorithms that can be utilized in the quantification of disease status, thus offering beneficial data to create personalized treatment. Furthermore, they are capable of measuring features in an imaging examination including intensity, shape, texture, wavelet, log-linear regression (LOG) features, among others, which enables prognostic or predictive non-invasive or imaging modalities to be built.[32] It is possible to use the platform for extraction Radiomics features from 2D and 3D images as well as binary masks on various imaging modalities. In this study, the overall number of periodontal defects included consisted of 58 in total. A random method was applied to separate the training and test datasets based on a 2:8 ratio, and there were 704 random seeds.

Image segmentation

Two senior radiologists with 10 and 5 years of experience reviewed all the images, and a radiologist who did not have access to the patients' clinical data manually delineated each of the disease lesions (VOI), and the senior radiologist reviewed each of the contours. If the discrepancy was $\geqslant 5\%$, the senior radiologist decided on the periodontal defect. Ultimately, 276 VOIs were segmented from 87 periapical images, which were employed in the subject analysis (Figure 2).

Feature extraction

Instead of developing a specific software tool for feature extraction, the image features are extracted using the web-based user interface provided by the Radiomics platform. These radiomic features were under the first-order, shape, and texture classifications. In particular, the texture features containing gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), and gray level size zone matrix (GLSZM) were used. Meanwhile, intensity and texture features can be calculated on the original image and derived images, obtained by applying several filters such exponential, logarithm, square, square root, and wavelet (wavelet-LHL, wavelet-L, wavelet-HLL, wavelet-LLH, wavelet-HLH, wavelet- HHH, wavelet-HHL, wavelet-LLL). Features comply with definitions as defined by the Imaging Biomarker Standardization Initiative (IBSI).[33] To reduce the dimensionality of features, variance threshold methods were used to gradually select the optimal features. A variance threshold was also applied (variance threshold = 0.8) to reduce the features.

The Radiomics platform was used to extract 1409 quantitative imaging features in total, which were assigned to three separate groups: Group 1 (first-order statistics) comprised 126 descriptors which quantitatively delineated the voxel intensity distribution in the image via simple and frequently utilized metrics. Group 2 (shape- and size-based features) included 14 3D features that denoted the region's size and shape. Group 3 contained 525 textual features that could quantify the regional heterogeneity, the calculation of which was made based on the gray level run length and co-occurrence texture matrices (texture features) (Table 1).

Feature qualification

Measurements were made of a significant amount of image features. Dimension reduction was conducted and task-specific features were chosen for the purpose of identifying which features were suitable. To lessen feature redundancy, methods applied when selecting features comprised variance threshold (variance threshold = 0.8), SelectKBest, as well as the least absolute shrinkage and selection operator (LASSO), which were employed for detecting whether significant differences existed between the low- and high-risk groups. Eigenvalues whose variance < 0.8 were excluded. The SelectKBest method was implemented with a p-value for the purpose of analyzing the correlation between the features and classification outcomes. Every feature that had p-values < 0.05 was utilized. The LASSO model's cost function was the L1 regularizer, the cross-validation error value was set at 5, while the maximum iteration number was 1000.

Machine learning analysis

Computer-generated random numbers were used to assign 80% of the VOIs to the training data set and 20% of VOIs to the test data set. Based on the selected features, there are several supervised learning classifiers available for classification analysis, which creates models that attempt to separate or predict the data with respect to an outcome or phenotype (for instance, patient outcome or response). In this study, the radiomics-based models were constructed with six classifiers, k-nearest neighbor (KNN), support vector

Table 1. Radiomic features selected for quantifying the heterogeneity differences.

Radiomic group	Associated filter	No. of features	Radiomic features
First-order statistics	None	126	Energy, total energy, entropy, minimum, 10 percentile, 90 percentile, maximum, mean, median, interquartile range, range, mean absolute deviation, robust mean absolute deviation, root mean square, standard deviation, skewness, kurtosis, variance
Shape	None	14	Volume, surface area, surface volume ratio, spherical disproportion, maximum 3D diameter, maximum 2D diameter column, maximum 2D diameter row, elongation
Texture features	GLCM	525	Autocorrelation, average intensity, cluster prominence, cluster shade, cluster tendency, contrast, difference average, difference entropy, difference variance, dissimilarity, entropy, sum average, sum entropy, sum variance, sum squares
Texture features	GLSZM		Large area emphasis, gray level non uniformity, size zone non uniformity, gray level variance, zone entropy, high gray level zone emphasis, small area high gray level emphasis, large area high gray level emphasis
Texture features	GLRLM		Gray level non uniformity, run length non uniformity, gray level variance, run entropy, high gray level run emphasis, short run high gray level emphasis, long run high level emphasis

machine (SVM), eXtreme Gradient Boosting (XGBoost), random forest (RF), logistic regression (LR) and decision tree(DT), and the test method was used to improve the effectiveness of the model. Furthermore, the best learning classifier will be selected for testing the periodontal defects.

Statistical analysis

The Radcloud platform was used in performing the statistical analysis of the data. From all the VOIs, 80% were assigned to the training dataset and 20% to the validation dataset according to random numbers generated via computer. The model was built using sic different classifiers, namely logistic regression (LR), random forest (RF), decision tree (DT), k-nearest neighbors (KNN), XGBoost, and support vector machine (SVM), which enabled the prediction of periodontal defects. Model performance evaluations considered specificity, sensitivity as well as receiver operating characteristic (ROC) curves. The point where the specificity plus sensitivity reached a maximum denoted the optimal cut-off value. Both the area under the curve (AUC) and accuracy of predictions were determined for the training and validation datasets. Subsequently, the classifier performance was evaluated using four separate indicators, namely P (precision = true positives/(true positives + false positives)), R (recall = true positives/(true positives + false negatives)), F1-score (F1-score = $P \times R \times 2/(P + R)$), support (total number in test set) was used for evaluating classifier performance.

$$P = Precision = True\ Positives/False\ Positives$$

$$R = Recall = True\ Positives/(True\ Positives + False\ Negatives)$$

$$F1Score = (P \times R \times 2)/(P + R)$$

Furthermore, the best learning classifiers were selected for comparison of the observer and machine learning models of periodontal defects evaluations. Calculation of the Kappa coefficients was performed to ascertain the intra- and inter-observer agreements. Kappa values were categorized as: $\kappa < 0.00$, no agreement; $\kappa = 0.00$–0.20, poor agreement; $\kappa = 0.21$–0.40, fair agreement; $\kappa = 0.41$–0.60, moderate agreement; $\kappa = 0.61$–0.80, good agreement; and $\kappa = 0.81$–1.00, very good agreement[34] between ML models and human observer evaluations. Calculation of the Kappa values was performed on the basis of these two categories.

The Mann-Whitney U test was utilized to test for statistical differences among the models' predictive radiomic features. Moreover, differences in clinical features among groups with respect to outcome parameters were tested using the univariate Fishers' exact test. IBM SPSS Statistics 21.0 (IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp) was used. If the p-value was less than 0.05, it was considered to be significant.

Results

From the total of 1409 total features, 411 features were selected utilizing the variance threshold technique (Figure 3), and the SelectKBest method was used to determine 10 features (Figure 4). Lastly, the Lasso algorithm was used, which defined four features that were optimal (Figure 5). The selected radiomic features were grouped into three, namely; Group 1 (first-order statistics) consisted of four descriptors that quantitatively delineate the distribution of voxel intensities within the image through commonly used and basic

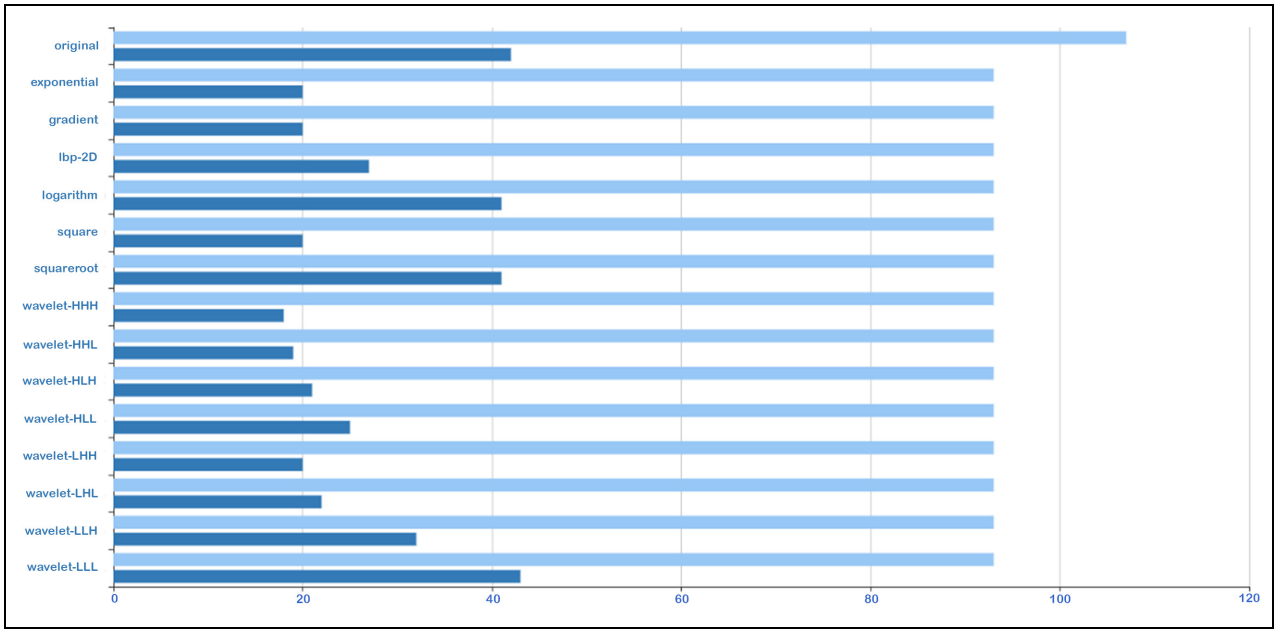


Figure 3. Variance threshold on feature selected. The variance threshold method was used to select radiomics features (variance threshold = 0.8), we selected 414 features from 1409 features.

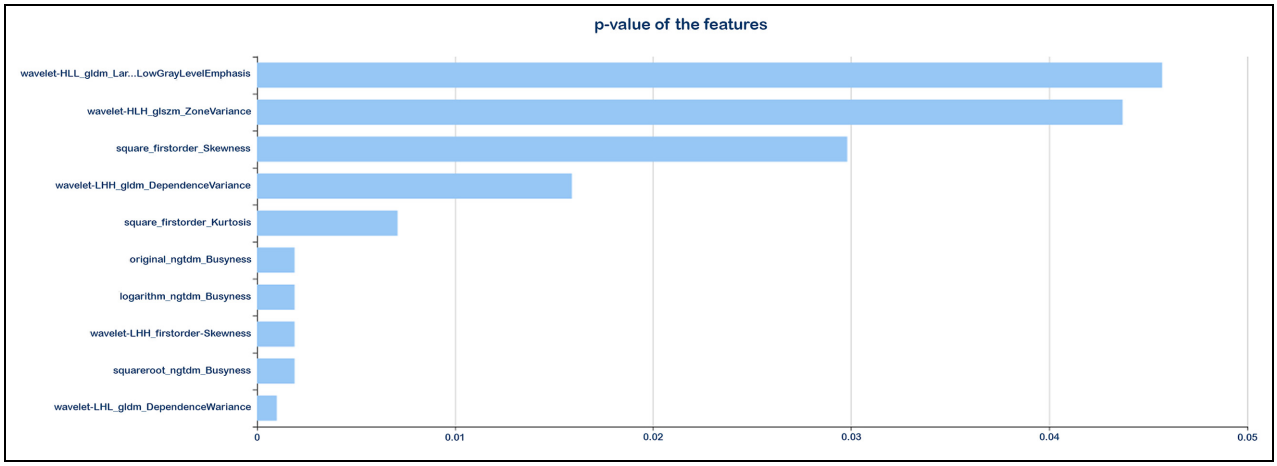


Figure 4. Select K-best method was used to further select the radiomics features; 10 features were selected.

metrics. Group 2 (shape- and size-based features) contained 8 two and three-dimensional features that reflect the shape and size of the region. Calculated from gray-level run-length and gray level co-occurrence texture matrices, 12 textural features that can quantify region heterogeneity differences.

Following an mRMR algorithm was applied. After this algorithm four features remained and were subjected to further selection by the LASSO algorithm.[35] Figure 6 depicts the confusion matrix after selecting radiomic features in the cohorts of groups, it was observed that these features had initially significant prognostic performance (CI $>$ 0.5) to differentiate periodontal defect exist or not (group 0–1). These four radiomic features were chosen for the purpose of identifying periodontal defects, namely, wavelet-HLH_glszm_ZoneVariance; original_ngtdm_Busyness; logarythm_ngtdm_Busyness; wavelet-LHH_firstorder_Skewness.

Tables 2 and 3 shows the AUC sensitivity and specificity of six classifiers in training and test set for periodontal defect detection in the SWM classifier were determined to be optimal method in training and test tests in terms of diagnostic accuracy according to the four indicators (Figure 7). Tables 4 and 5 summaries the four indicators (precision, recall, F1-score, support) for six classifiers for detecting of periodontal defect.

Table 6 presents the kappa values for both observer and ML methods to detect periodontal defects utilizing the optimal MLP learning classifier (SWM). Compared

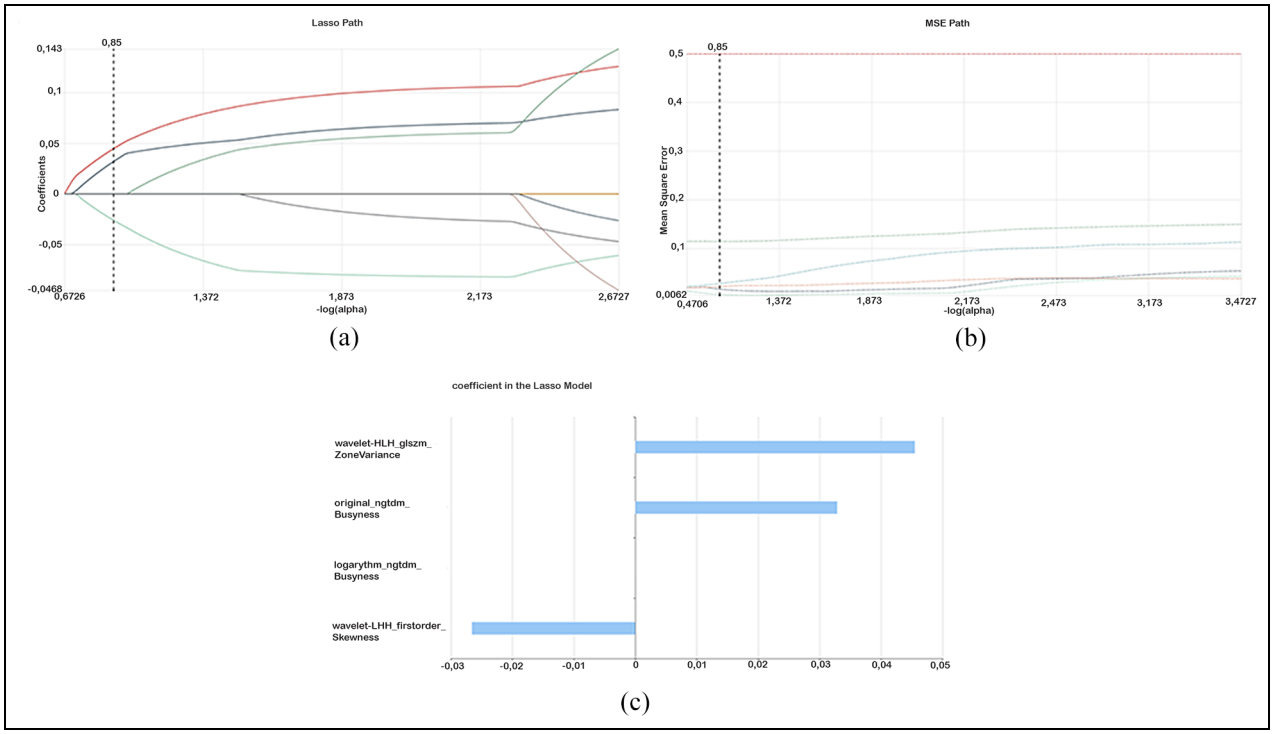


Figure 5. Lasso algorithm for feature selection: (a) Lasso path, (b) MSE path, and (c) coefficients in the Lasso model. The Lasso model was used to select four features that correspond to the optimal alpha value. Four features were selected.

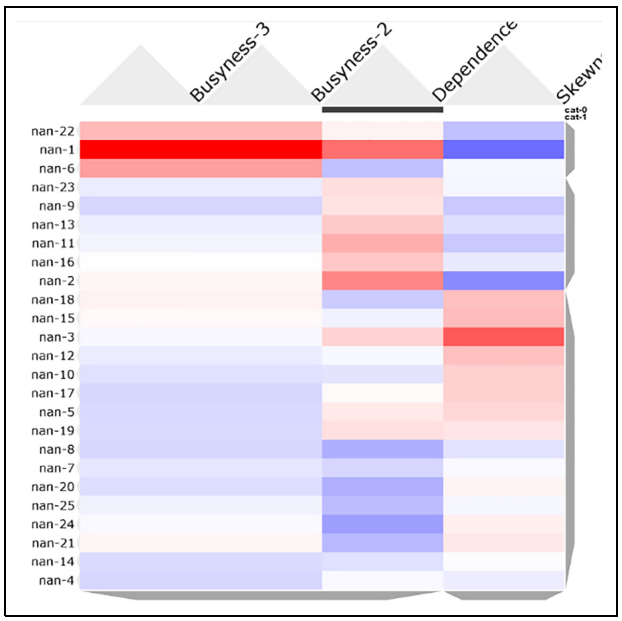


Figure 6. Depicts the confusion matrix after selecting radiomic features in the cohorts of groups.

with the gold standard, the ability of both human observers and the ML to detect periodontal defects was significantly different ($p < 0.05$). However, ML and Human observers exhibited similar performance in detecting periodontal defects with no significant differences ($p > 0.05$).

Discussion

In periodontics, a suitable treatment plan can only be determined for a patient if they are correctly diagnosed, thus ensuring that their health can be restored. While the knowledge of the specialist has particular importance for diagnosing the condition and subsequent treatment, they can be exposed to significant risk of error based on multiple factors.[12] When diagnosing periodontal diseases, in addition to important findings such as bleeding, tooth hypermobility, and gingival features on probing used in clinical examination, it is also important to detect existing bone levels, bone loss, and intraosseous defects with dental radiographs.[36]

Each radiographic method has its advantages and disadvantages, and clinical findings and the benefit to the patient are important in the selection of the technique to be used. In the review published by Kumar et al. in 2014, it was reported that intraoral radiography is superior to panoramic radiographs in detecting bone defects. Although CBCT is considered a useful complement, it has been stated that radiation dose and clinical benefit should be taken into account.[37] Existing evidence reinforces the assertion that 2D full-mouth radiographic series be used in conjunction with clinical probing parameters as the benchmark approach for comprehensively evaluating periodontal structures.

2D radiographs provide the greatest amount of diagnostic details and are frequently employed for evaluating periodontal conditions. However, the internal morphology, depth of crater-like defects, or the extent

Table 2. ROC results with KNN, SVM, XGBoost, RF, LR, and DT classifiers on the training set.

Classifiers	Category	AUC	95% CI	Sensitivity	Specificity
KNN	0	0.922	0.82–1.00	0.81	1
	1	0.922	0.82–1.00	1.00	0.81
SVM	0	1	1.00–1.00	1.00	1
	1	1	1.00–1.00	1.00	1
XGBoost	0	0.938	0.85–1.00	0.88	1
	1	0.938	0.85–1.00	1.00	0.88
RF	0	1	1.00–1.00	1.00	1
	1	1	1.00–1.00	1.00	1
LR	0	0.938	0.85–1.00	0.88	1
	1	0.938	0.85–1.00	1.00	0.88
DT	0	1	1.00–1.00	1.00	1
	1	1	1.00–1.00	1.00	1

Category 1: periodontal defect is exist; category 0: no periodontal defect.

Table 3. ROC results with six classifiers on the test set.

Classifiers	Category	AUC	95% CI	Sensitivity	Specificity
KNN	0	0.25	−0.01–0.51	0.50	0
	1	0.25	−0.01–0.51	0.00	0.5
SVM	0	0.75	0.75–0.75	0.00	1
	1	0.75	0.75–0.75	1.00	0
XGBoost	0	0.375	0.38–0.38	1.00	0
	1	0.375	0.38–0.38	0.00	1
RF	0	0.25	0.25–0.25	1.00	0
	1	0.25	0.25–0.25	0.00	1
LR	0	0.25	0.25–0.25	1.00	0
	1	0.25	0.25–0.25	0.00	1
DT	0	0.5	0.50–0.50	1.00	0
	1	0.5	0.50–0.50	0.00	1

Category 1: periodontal defect is exist; category 0: no periodontal defect.

of involvement on the facial and lingual surfaces are not revealed by 2D radiographs. Due to its high sensitivity, DSR is used to detect minimal changes in the dento-alveolar structure but requires a series of radiographic images, which means more radiation doses. In addition to being versatile images that can support the diagnosis of periodontal defects, 2D radiographs can be used to gather information to create artificial intelligence based on computer vision through various processes such as ML. Automatic detection of periodontal defects can save time and reduce human errors caused by fatigue and workload.

There has been increased usage of deep learning in computer vision for analyzing medical images in recent decades. In particular, the compatibility of AI with image processing has highlighted the studies on radiology. Diagnosis of caries, the numbering of teeth, detection of periodontal diseases, determination of anatomical structures, examination of cephalometric analysis, and diagnosis of some diseases such as osteoporosis, which manifests in the jaws, are among the current topics studied.[38] In these studies, analysis was conducted on a broad range of radiographic images including periapical, panoramic, bitewing cephalometric, CT, and CBCT. To the best of the authors' knowledge, no previous studies in the literature have focused on determining periodontal defects such as dehiscence, fenestration, or early furcation defects, and therefore, this represents the first attempt to study this topic.

In this research, the machine learning model was used to detect periodontal defects from periapical radiographic images. The study findings indicated that success close to the gold standard was achieved in detecting periodontal defects through artificial intelligence. Bone defects on periapical radiographs could be detected directly by reducing possible errors and eliminating doctor-related conditions with AI. Our study outcomes are analogous to those of other researchers who evaluated periodontal bone loss with artificial intelligence applications using panoramic radiographs and AI gives similar or superior results to dentists.

Different machine learning strategies, such as k-nearest neighbors, support vector machines, or random forest decision trees, can be applied to construct the

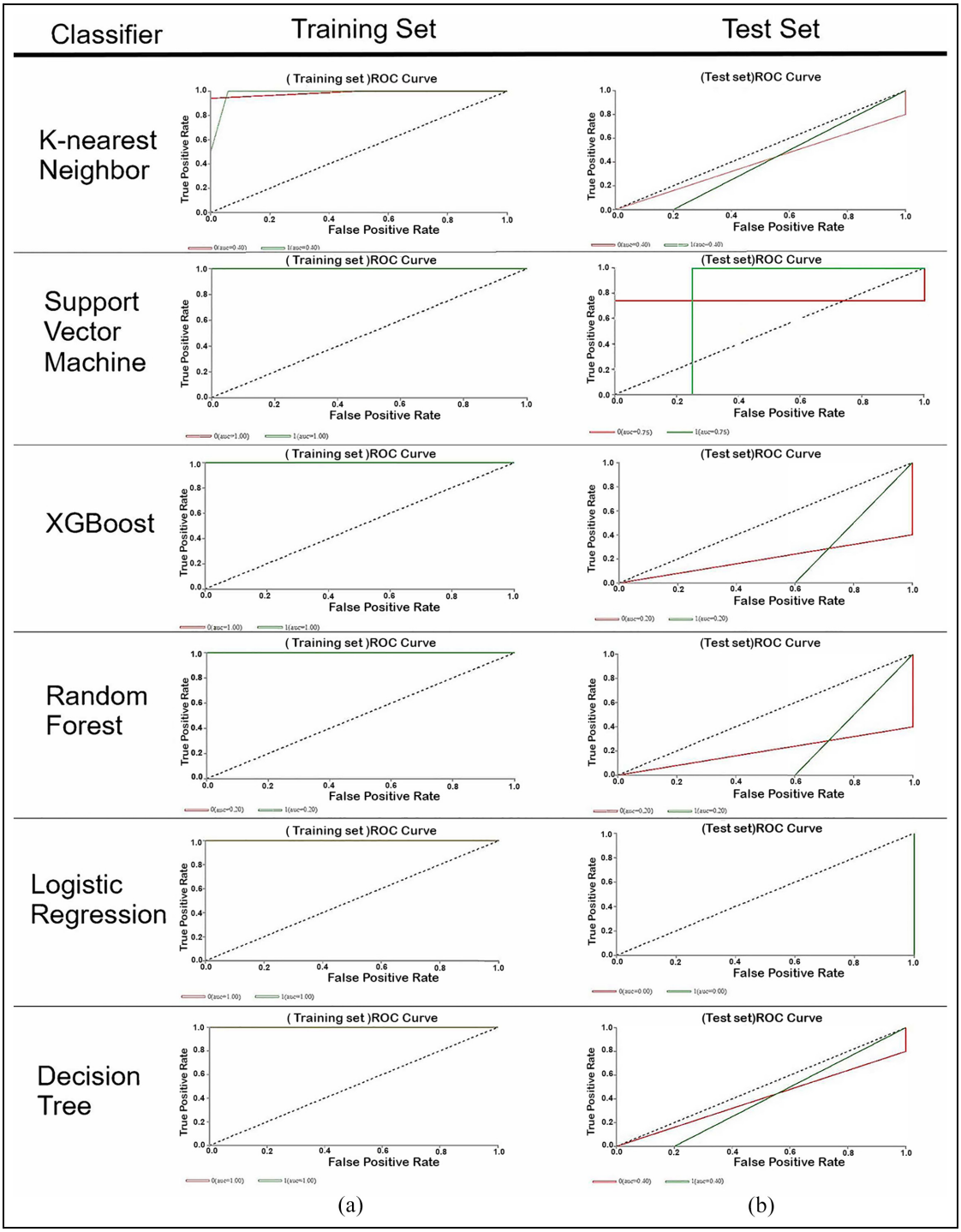


Figure 7. ROC curves of machine learning methods for classification. Green indicates non-defect red indicates defects. (a) ROC curve of the training dataset, (b) ROC curve of the test dataset.

mapping of a given training set and a given set of features. During training, the parameters that define the mapping whose representation depends on the chosen learning strategy, are iteratively refined such that estimation performance is maximized on the training set itself, together with this evaluation the difference between the given ground truth for each image and in the training set can be evaluated.[39] Support vector machine (SVM) is a computer algorithm that learns by pattern to assign labels to objects and needs long training time for large datasets.[28,40] SVM algorithms set up an imaginary high-dimensional space and place the samples in the appropriate locations according to their features.

Several previous studies indicated the use of AI for periodontal pathology diagnosis. Chang et al.[41] designed an automated technique to stage periodontitis on panoramic dental X-rays through the use of the

Table 4. The results of four indicators – precision, recall, F1-score, support on the training set.

	Indicators	KNN	SVM	XGBoost	RF	LR	DT
0	Precision	1.00	1.00	1.00	1.00	1.00	0.00
	Recall	0.81	1.00	0.88	1.00	0.88	0.00
	F1-score	0.90	1.00	0.93	1.00	0.93	0.00
	Support	16	16	16	16	16	16
1	Precision	0.40	1.00	0.50	1.00	0.50	0.00
	Recall	1.00	1.00	1.00	1.00	1.00	0.00
	F1-score	0.57	1.00	0.67	1.00	0.67	0.00
	Support	2	2	2	2	2	2

Category 1: periodontal defect is exist; category 0: no periodontal defect.

Table 5. The results of four indicators – precision, recall, F1-score, support in testset.

	Indicators	KNN	SVM	XGBoost	RF	LR	DT
0	Precision	0.67	0.80	0.80	0.80	0.80	0.00
	Recall	0.50	1.00	1.00	1.00	1.00	0.00
	F1-score	0.57	0.89	0.89	0.89	0.89	0.00
	Support	4	4	4	4	4	4
1	Precision	0.00	0.20	0.00	0.00	0.00	0.00
	Recall	0.00	1.00	0.00	0.00	0.00	0.00
	F1-score	0.00	0.33	0.00	0.00	0.00	0.00
	Support	1	1	1	1	1	1

Category 1: periodontal defect is exist; category 0: no periodontal defect.

Table 6. The kappa values for the observer and ML methods to detect periodontal defects using the best MLP learning classifier (SWM).

	Kappa	*p*	Sensitivity	PPV	Specificity	NPV
Gold standard – first reading*	0.752	0.0001	85	98.1	96.3	74.3
Gold standard – second reading*	0.692	0.0001	78.3	100	100	67.5
Gold standard – ML*	0.709	0.0001	80	96	92.6	67.6
ML – first reading	0.716	0.0001	90	86.5	81.1	85.7
ML – second reading	0.651	0.0001	82	87.2	83.8	77.5
First reading – second reading	0.883	0.0001	90.4	100	100	87.5

ML: machine learning; NPV: negative predictive value; PPV: positive predictive value.
Statistical significance $p < 0.05$.
*Indicated without significant difference among one reading – second reading and ML detection for periodontal defects.

deep learning hybrid technique. In this study, in which deep learning was used for the detection of radiographic bone level (or CEJ level) as a basic structure for the whole jaw on panoramic X-rays, it was found that this approach was highly accurate and reliable for automatically diagnosing bone loss due to periodontal disease as well as staging the disease. The authors suggested that the developed method could help dentists with the diagnosis as well as the systematic and precise monitoring of periodontitis on panoramic X-rays, diagnose, and systematically and precisely monitor periodontitis on panoramic radiographs; thus, it could provide a significant improvement in dentists' ability to diagnose and treat periodontitis.

Kim et al.[42] proposed a deep learning-based method called DeNTNet to develop an automated diagnostic support system detecting bone loss due to periodontitis capable of panoramic dental X-rays. It has been suggested that a superior PBL detection performance can be obtained with the model proposed in the study than that of dentists.

Similarly, Krois et al.[43] applied deep convolutional neural networks (CNNs) for the detection of periodontal bone loss (PBL) on panoramic dental X-rays. It was reported that after being trained on a small number of radiographic image segments, a CNN showed a largely similar performance compared with dentists in terms of the assessment of PBL on panoramic X-rays. In line with

previous studies, high levels of accuracy and sensitivity were recorded in our research. However, the ability to compare with other studies is limited due to the different methods used as Krois et al.[43] used deep learning CNN, whereas our method is based on machine learning.

In this study, the highest AUC has been achieved with MLP classifier (SWM) machine learning method than the other learning classifiersfor periodontal defect detection in the SWM, the classifiers were determined to be optimal methods in training and test tests in terms of diagnostic accuracy according to the indicators. Furthermore, three radiomic features were chosen for the purpose of identifying periodontal defects, namely, wavelet-HLH_glszm_ZoneVariance; original_ngtdm_Busyness; logarythm_ngtdm_Busyness; wavelet-LHH_firstorder_Skewness.

Our study had certain limitations. The primary limitation of our study is that a relatively small data set was used. Even though the volume of training data used in our study for deep learning was comparatively limited, the level of performance achieved in detecting periodontal defects was still high. Although 2D radiographs have relatively low radiation doses, provide good diagnostic data, and are frequently utilized for evaluating periodontal disease, they have limitations in detecting alveolar bone defects. As it is highly sensitive and precise, DSR is used for diagnosing and following-up on the loss of alveolar bone caused by periodontal disease as well as evaluating treatment outcomes. However, DSR requires serial radiographic images of the anatomical structure of interest. This causes patients to be exposed to more radiation. Detection of periodontal defects with artificial intelligence eliminates this handicap and allows the defect to be determined with a single periapical radiograph.

Another limitation of our study radiographic images obtained from defects prepared in dry human mandibles under standard conditions were evaluated. Considering the necessity for data standardization in the dentistry field, as well as the importance of standardization of data set structure, this is the superiority of our study. Additionally, in this study, supervised machine learning classifiers was performed using a prelabeled periapical radiographic dataset. We have also shown that the results have diagnostic and predictive accuracy similar to those obtained by an experienced periodontist. The future research should also be done in CBCT images in 3D nature to figure out how this AI ML modeling can be beneficial in terms of CBCT imaging.

Conclusions

In this research, a machine learning model involving SVM analysis of digital dental images capable of classifying periodontal defects has been proposed. Additionally, this study has shown that combining particular radiomic features with image variables can enable the prediction of periodontal defects. The proposed AI system can be utilized in the support of clinical practitioners and ultimately replace evaluation by human observers and improve future performance.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Fatma Karacaoglu https://orcid.org/0000-0003-0122-2038

References

1. Ridgeway EE. Periodontal disease: diagnosis and management. *J Am Acad Nurse Pract* 2000; 12: 79–84.
2. Pihlstrom BL, Michalowicz BS and Johnson NW. Periodontal diseases. *Lancet* 2005; 366(9499): 1809–1820.
3. Vandenberghe B, Jacobs R and Yang J. Detection of periodontal bone loss using digital intraoral and cone beam computed tomography images: an in vitro assessment of bony and/or infrabony defects. *Dentomaxillofac Radiol* 2008; 37(5): 252–260.
4. Kolsuz ME and Paksoy CS. Periodontal kemik defektlerinin varlığı ve bukkal-lingual lokalizasyonlarının belirlenmesinde farklı intraoral görüntüleme yöntemleri ile dijital subtraksiyon tekniğinin etkinliklerinin karşılaştırılması. *Turkiye Klinikleri Dis Hekimligi Bilimler iDergisi* 2016; 22(1): 21–28.
5. Gröndahl HG and Gröndahl K. Subtraction radiography for the diagnosis of periodontal bone lesions. *Oral Surg Oral Med Oral Pathol* 1983; 55(2): 208–213.
6. Christgau M, Wenzel A, Hiller KA, et al. Quantitative digital subtraction radiography for assessment of bone density changes following periodontal guided tissue regeneration. *Dentomaxillofac Radiol* 1996; 25(1): 25–33.
7. Ellwood RP, Davies RM and Worthington HV. Evaluation of a dental subtraction radiography system. *J Periodontal Res* 1997; 32(2): 241–248.
8. Mol A and Dunn SM. The performance of projective standardization for digital subtraction radiography. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2003; 96(3): 373–382.
9. Lee SS, Huh YJ, Kim KY, et al. Development and evaluation of digital subtraction radiography computer program. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2004; 98(4): 471–475.
10. de Faria Vasconcelos K, Evangelista KM, Rodrigues CD, et al. Detection of periodontal bone loss using cone beam CT and intraoral radiography. *Dentomaxillofac Radiol* 2012; 41(1): 64–69.
11. Bayat S, Talaeipour AR and Sarlati F. Detection of simulated periodontal defects using cone-beam CT and digital

intraoral radiography. *Dentomaxillofac Radiol* 2016; 45(6): 20160030.
12. Machoy M, Szyszka-Sommerfeld L, Vegh A, et al. The ways of using machine learning in dentistry. *Adv Clin Exp Med* 2020; 29(3): 375–384.
13. Reyes LT, Knorst JK, Ortiz FR, et al. Machine learning in the diagnosis and prognostic prediction of dental caries: a systematic review. *Caries Res* 2022; 56: 161–170.
14. Alharbi MT and Almutiq MM. Prediction of dental implants using machine learning algorithms. *J Healthc Eng* 2022; 2022: 7307675.
15. Hasuike A, Watanabe T, Wakuda S, et al. Machine learning in predicting tooth loss: a systematic review and risk of bias assessment. *J Pers Med* 2022; 12: 1682–1693.
16. Schwendicke F, Samek W and Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res* 2020; 99(7): 769–774.
17. Xing F and Yang L. Machine learning and its application in microscopic image analysis. In: G Wu, D Shen and MR Sabuncu (eds) *Machine learning and medical imaging*. Amsterdam: Elsevier Inc., 2016, pp.97–127.
18. Badrick T, Banfi G, Bietenbeck A, et al. Machine learning for clinical chemists. *Clin Chem* 2019; 65(11): 1350–1356.
19. Sultan AS, Elgharib MA, Tavares T, et al. The use of artificial intelligence, machine learning, and deep learning in oncologic histopathology. *J Oral Pathol Med* 2020; 49: 849–856.
20. Chartrand G, Cheng PM, Vorontsov E, et al. Deep learning: a primer for radiologists. *Radiographics* 2017; 37: 2113–2131.
21. Jones LD, Golan D, Hanna SA, et al. Artificial intelligence, machine learning and the evolution of healthcare. *Bone Joint Res* 2018; 7: 223–225.
22. Burges CJC. Some notes on applied mathematics for machine learning. In: O Bousquet, U Luxburg and G Rätsch (eds) *Advanced Lectures on machine learning*. Heidelberg: Springer-Verlag Berlin, 2003, pp.21–40.
23. Gillies RJ, Kinahan PE and Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016; 278: 563–577.
24. Tan Y, Schwartz LH and Zhao B. Segmentation of lung lesions on CT scans using watershed, active contours, and Markov random field. *Med Phys* 2013; 40(4): 043502.
25. Panth KM, Leijenaar RTH, Carvalho S, et al. Is there a causal relationship between genetic changes and radiomics-based image features? An in vivo preclinical experiment with doxycycline inducible GADD34 tumor cells. *Radiother Oncol* 2015; 116: 462–466.
26. Rizzo S, Petrella F, Buscarino V, et al. CT radiogenomic characterization of EGFR, K-RAS, and ALK mutations in non-small cell lung cancer. *Eur Radiol* 2016; 26: 32–42.
27. Salvador-Meneses J, Ruiz-Chavez Z and Garcia-Rodriguez J. Compressed kNN: K-nearest neighbors with data compression. *Entropy* 2019; 21: 234.
28. Noble WS. What is a support vector machine? *Nat Biotechnol* 2006; 24: 1565.
29. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; 48(4): 441–446.
30. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imaging* 2012; 30(9): 1234–1248.
31. Wang YY, Zhang T, Li SW, et al. Mapping p53 mutations in low-grade glioma: a voxel-based neuroimaging analysis. *Am J Neuroradiol* 2015; 36(1): 70–76.
32. Burges CJC. Some notes on applied mathematics for machine learning. In: O Bousquet, U Luxburg and G Rätsch (eds) *Advanced lectures on machine learning*. Berlin Heidelberg: Springer-Verlag, 2003, pp.21–40.
33. Zwanenburg A, Leger S, Vallières M, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 2020; 295: 328–338.
34. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20: 37–46.
35. De Jay N, Papillon-Cavanagh S, Olsen C, et al. MRMRe: an R package for parallelized mRMR ensemble feature selection. *Bioinformatics* 2013; 29: 2365–2368.
36. Corbet EF, Ho DKL and Lai SML. Radiographs in periodontal disease diagnosis and management. *Aust Dent J* 2009; 54(1 Suppl): S27.
37. Kirandeep Arora VK. Different radiographic modalities used for detection of common periodontal and periapical lesions encountered in routine dental practice. *J Oral Hyg Health* 2014; 2: 5.
38. Hwang JJ, Jung YH, Cho BH, et al. An overview of deep learning in the field of dentistry. *Imaging Sci Dent* 2019; 49(1): 1–7.
39. Prince SJD. *Computervision: models, learning, and inference*. 1st ed. Cambridge: Cambridge University Press, 2012.
40. Nalepa J and Kawulok M. Selecting training sets for support vector machines: a review. *Artif Intell Rev* 2019; 52: 857–900.
41. Chang HJ, Lee SJ, Yong TH, et al. Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis. *Sci Rep* 2020; 10(1): 8.
42. Kim J, Lee H-S, Song IS, et al. DeNTNet: deep neural transfer network for the detection of periodontal bone loss using panoramic dental radiographs. *Sci Rep* 2019; 9(1): 9.
43. Krois J, Ekert T, Meinhold L, et al. Deep learning for the radiographic detection of periodontal bone loss. *Sci Rep* 2019; 9(1): 6.