Nobel Med 2024; 20(2): 112-118

EVALUATION OF THE PERFORMANCE OF MACHINE LEARNING IN CLASSIFICATION OF IMAGES WITH OR WITHOUT MISSING TEETH IN PANORAMIC RADIOGRAPHS

Kübra Törenek Ağırman, Kübra Başaran Aslan

[1]*Atatürk University, Faculty of Dentistry, Department of Dentomaxillofacial Radiology, Erzurum, Türkiye*

ABSTRACT

Objective: *This study aims to evaluate the functionality and usability of machine learning (ML) in classifying missing teeth in panoramic radiography.*

Material and Method: *In this study, of 1000 anonymous panoramic radiographs archived for the classification of missing teeth, 500 contained missing teeth, while the other 500 did not contain missing teeth. 700 of the images are reserved for training and 300 for testing. Principal component analysis (PCA) was used to extract features from panoramic radiographs. Six different classification model algorithms (Support Vector Machines (SVM), Random Forest Classifier, Logistic Regression, KNeighbors Classifier, Decision Tree Classifier, and Gaussian NB) were used for missing/complete tooth classification on the created data set. The performance of these models was evaluated.*

Results: *Among the classification models included in the study, the accuracy scores of SVM were found to be higher than other algorithms, with 98.14% in the training data set and 81.67% in the test data set.*

Conclusion: *The selection of the appropriate machine learning model is very important to ensure accurate and reliable diagnosis in the field of medical image classification. SVM is a very successful method in classifying multidimensional data.*

Keywords: *Missing teeth, machine learning, panoramic radiography.*

C **CORRESPONDING AUTHOR: Kübra Başaran Aslan** Department of Dentomaxillofacial Radiology, Faculty of Dentistry, Ataturk University, 25240, Erzurum, Türkiye dtkubrabasaran@gmail.com

ORCID **KTA https:**//orcid.org/0000-0001-7200-3436 ORCID **KBA https:**//orcid.org/0000-0002-7705-2760

DELIVERING DATE: 08 / 08 / 2023 • **ACCEPTED DATE:** 13 / 11 / 2023

PANORAMİK RADYOGRAFİLERDE EKSİK DİŞLİ VE DİŞSİZ GÖRÜNTÜLERİN SINIFLANDIRILMASINDA MAKİNE ÖĞRENMESİ PERFORMANSININ DEĞERLENDİRİLMESİ

ÖZET

Amaç: Bu çalışma, panoramik radyografide eksik dişlerin sınıflandırılmasında makine öğreniminin (ML) işlevselliğini ve kullanılabilirliğini değerlendirmeyi amaçlamaktadır.

Materyal ve Metot: Bu çalışmada, eksik dişlerin sınıflandırılması için arşivlenen 1000 anonim panoramik radyografinin 500'ü eksik dişleri içerirken, diğer 500'ü eksik dişleri içermiyordu. Görüntülerin 700'ü eğitim, 300'ü ise test için ayrılmıştır. Panoramik radyografilerden özellik çıkarmak için temel bileşen analizi (PCA) kullanıldı. Oluşturulan veri seti üzerinde eksik/tam diş sınıflandırması için olmak üzere altı farklı sınıflandırma modeli algoritması (Destek Vektör Makineleri (SVM), Rastgele Karar Ormanı, Lojistik Regresyon, K-En Yakın Komşu, Karar Ağacı ve Gaussian Naive Bayes sınıflandırma modeli) kullanıldı. Bu modellerin performansı değerlendirildi.

Bulgular: Çalışmaya dahil edilen sınıflandırma modelleri içerisinde SVM'nin doğruluk puanları eğitim veri setinde %98,14, test veri setinde ise %81,67 ile diğer algoritmalara göre yüksek bulundu.

Sonuç: Tıbbi görüntü sınıflandırma alanında doğru ve güvenilir teşhisin sağlanması için uygun makine öğrenmesi modelinin seçimi oldukça önemlidir. SVM, çok boyutlu verilerin sınıflandırılmasında oldukça başarılı bir yöntemdir.

Anahtar kelimeler: Eksik dişler, makine öğrenimi, panoramik radyografi.

INTRODUCTION

Detection and elimination of missing teeth are of great importance in increasing the quality of life of patients. However, even clinically experienced dentists may abduct congenitally missing or extracted teeth.[1] Dentists generally prefer panoramic radiographs to evaluate the jaw and dental tissues together in routine examinations.[2]

Machine learning (ML) is the use of mathematical models to help people learn without the need for clear commands from a computer to decide what to do and it can be considered a subset of artificial intelligence (AI). ML typically begins with an algorithm system extracting features from images to be used in the prediction or diagnosis of interest. It then identifies the best combination of these image features to classify the image or calculate some measure for the given image region.[3] Deep Learning (DL), which we have heard frequently recently; is a complex subset of ML built on neural networks. One of the most basic features of artificial neural networks is that they consist of layers with multiple inputs and outputs.[4] The latest core model of artificial neural networks has been convolutional neural networks (CNNs).[5]

The machine learning methods employed both linear and nonlinear dynamics for data classification, building upon the concepts of linear and nonlinear cellular automata approaches utilized in the realm of pattern recognition.[6] While ML provides the opportunity to work on smaller datasets than DL, DL is more suitable for processing complex data. AI studies, which have recently increased in the health sector, are fascinating for physicians in terms of reducing the workload of physicians and preventing neglect.[7]

This study aims to evaluate the performance and usability of ML in the classification of missing teeth in panoramic radiography.

MATERIAL AND METHOD

Radiographic Dataset

A total of 1000 anonymized panoramic radiographs were gathered from our clinic image database stored between January 2020 and April 2020. This study was approved by the Atatürk University Faculty of Dentistry Research Ethics Committee and all stages were carried out as declared in the Helsinki Declaration guidelines. (decision no: 02/2023-14)

In this study, a total of 1000 radiology images (700 of the selected images randomly are reserved for training and 300 for testing) with and without 500 missing teeth, which were re-checked and confirmed by oral and maxillofacial radiology experts, were used. Images were selected from patients who had completed permanent dentition, and wisdom teeth were not taken into account when evaluating missing teeth.

Panoramic radiographs of patients with fixed prosthetic restorations and implants were also not included in the study and no distinction was made between missing teeth and congenitally missing teeth.

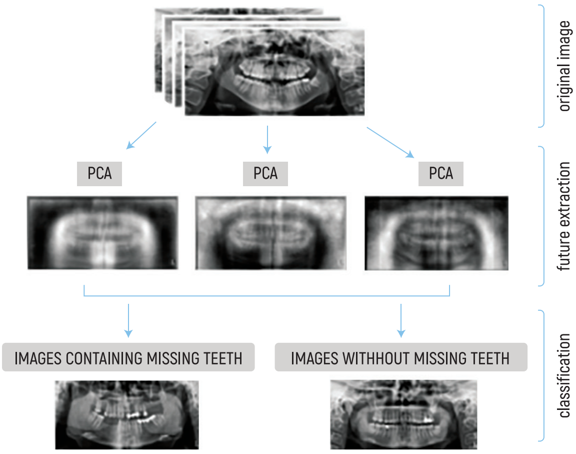

Figure 1. General procedure for hybrid framework of principal component analysis (PCA) approach of machine learning to classify the presence of missing teeth

Feature Extraction from Radiographic Images

Raw data are not directly input into ML models. First of all, preprocesses should be performed and features should be extracted from the data. In this study, resizing and scaling of pixel values were preprocessed on radiology images.

Unsupervised ML methods are used for dimensionality reduction or feature extraction. The most commonly used method among these is the principal component analysis (PCA) method. In this study, PCA was also used to extract features from radiographs and reduce dimensionality.

PCA uses an orthogonal transformation to transform a set of related vectors into an unrelated set of vectors. The important information contained in vectors, their lengths, and the angle between them, do not change. The information contained in vectors is defined by matrix multiplication and preserved by orthogonal transformation.

PCA creates a feature vector as follows: The first component was created to account for as much of the variability in the data as possible. Subsequent components are then created to contain varying amounts.[8] In general, a small portion of the components extracted with PCA represents the majority of the extracted feature vector.

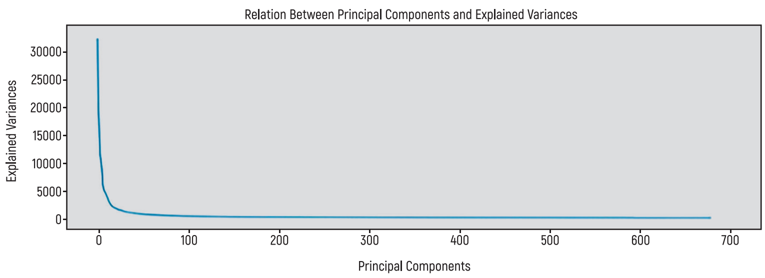

Figure 2. Trained PCA's Principal Components and Explained Variance Relationship.

A schematic of the general procedure for the hybrid framework of the PCA machine learning approach to classify the presence of missing teeth is given in Figure 1. In this study, a W projection matrix was created using the training data set to apply PCA to the training and test data sets. The following steps were performed to create the projection matrix from the training dataset:

1. All radiology images in the dataset were reduced to 358x739.
2. 358x739 image flattened to 264562 size vector.
3. 700 vectors (samples) are reserved for training.
4. A 700x264562 covariance matrix was created. Eigenvalues and eigenvectors of the covariance matrix of size 700x264562 were calculated.
5. 120 eigenvectors were selected, each with a length of 264562 and corresponding to 120 eigenvalues.
6. A projection matrix of size 264562x120 was created from 120 selected eigenvectors, each with a length of 264562. Thus, the size of the 264562 feature vectors can be reduced to 120. The number of basic components to be used was determined by examining the graph in Figure 2.

It seems that the variances explained above do not make much difference after 100. In the experimental study, it was evaluated that 120 basic components would be sufficient. The eigenimages obtained from the main components are listed in Figure 3.

The use of PCA in dimensionality reduction is achieved by a projection matrix W with dimensions dxk. The input vector of length d is multiplied by the projection matrix W and reduced to dimension k. Size reduction is performed as follows:[9]

x= x_2,…,x_d,x∈R^d

In Equation 1 above, x is the feature vector. The transformation of the feature vector is performed using the projection matrix W as follows:

xW=zW∈R^dxk

The z vector obtained from the equation will be as follows:

z= z_3,…,z_k,z∈R^k

Obtaining the projection matrix W in the above equation would be as follows using the existing data set:

- Since PCA is sensitive to scaling, the dataset is scaled first.
- The covariance matrix is created from the data set.
- Eigenvectors and eigenvalues are obtained from the covariance matrix.
- Eigenvalues are sorted in ascending order.

- Select the k eigenvectors, each of length d, corresponding to the k eigenvalues.
- W projection matrix is created from the selected k eigenvectors.

Classification of Data

The classification performances of support vector machine (SVM), Random Forest Classifier, Logistic Regression, KNeighbors Classifier, Decision Tree Classifier, and Gaussian NB, which are frequently preferred classification algorithms in machine learning, were evaluated for missing teeth. Since it is more successful in classifying large-sized data, SVM was taken as a guide and detailed.

Firstly, the SVM classification algorithm was used for missing/complete tooth recognition of feature vectors extracted from images using PCA. The optimization goal in SVM is to maximize margin. The margin is defined as the distance between the separation hyperplane (decision boundary) and the training samples closest to this hyperplane, called support vectors.[10] The SVM classifier is first trained on the defeated train dataset samples. The performance of the trained model was evaluated on training and testing datasets.

Let samples containing two classes with N instances are defined as follows:

$$X=\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n) \forall x_i \in R^d, y \in \{+1,-1\}\}$$

The size of the feature vector is represented by d and the classes to which it may belong are represented by +1 and -1. The hyperplane that separates the two classes together is considered as the decision function and is defined as follows:

$$f(x)=w^T x_i+b$$

The above decision function takes as its input the feature vector x of length d and does dot product with the weight vector w^T. The scalar result is summed up with the bias value. The result of the function is evaluated as follows:

$$\mathrm{sign}(f(x))=\begin{cases} +,\wedge f(x)>0 \\ -,\wedge f(x)<0 \end{cases}$$

The sign function above takes the result of the decision function as input. If the result is greater than 0, it returns to the positive class, otherwise it returns to the negative class.

Similarly, accuracy, precision, recall and F1-score values were obtained for each classification model.

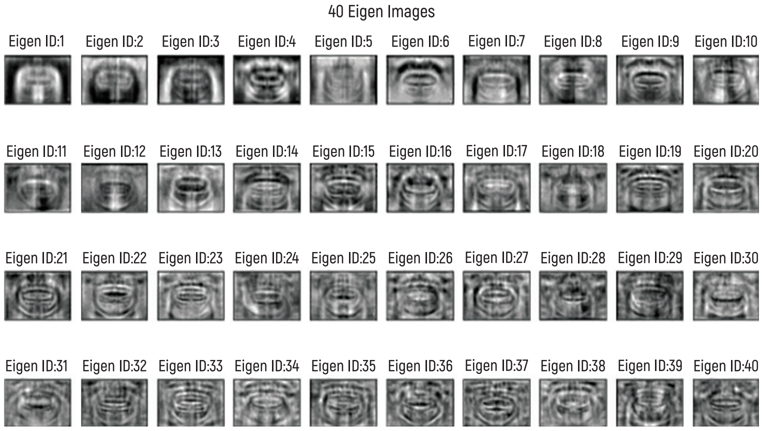

Figure 3. 40 Eigen Images derived from 40 principal components

RESULTS

With the selected classification model SVM, the success of the training model was determined to be 98.14%, and the success of the images used for testing was determined to be quite high at 81.67%. The Roc curve in the training and test data sets of the model is given in Figure 4, and the confusion matrix results are given in Figure 5. While SVM, Logistic Regression, and Gaussian NB exhibited good testing accuracy, models

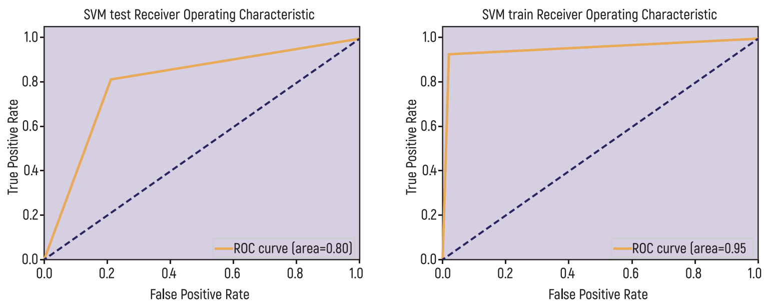

Figure 4. The Roc curve in the training and test data sets of the model

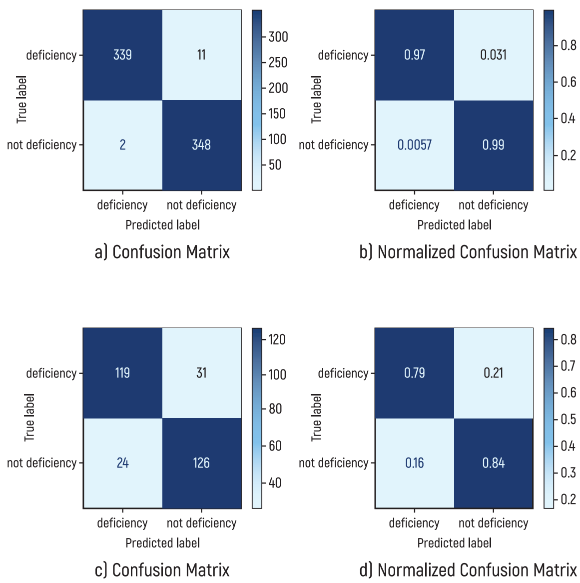

Figure 5. Confusion Matrices of the model for the train dataset (a, b) and the test dataset (c, d)

Table 1. Accuracy values of different classification models		
Model	**Train Accuracy**	**Test Accuracy**
SVM	98.14	81.67
Random Forest Classifier	99.86	74.00
Logistic Regression	88.11	81.00
KNeighbors Classifier	81.09	61.67
Decision Tree Classifier	99.86	65.33
Gaussian NB	83.52	70.00

SVM: Support vector machine

such as Random Forest Classifier and Decision Tree Classifier appeared to be overfitting, as evidenced by the large disparity between training and testing accuracies. The KNeighbors Classifier showed limited generalization ability. Accuracy, precision, recall and F1-score values on the test and training models of six different classifiers are presented in detail in Table 1 and Table 2.

DISCUSSION

In dental practice, x-ray radiography (periapical, panoramic radiographs), and cone-beam computed tomography (CBCT) are the most commonly used imaging modalities for tooth numbering, identification, diagnosis, and determination of treatment options.[11] Currently, studies have evaluated the potential accuracy of AI approaches in interpreting medical images such as X-rays radiography, computed tomography (CT), CBCT, ultrasonography (US), magnetic resonance imaging (MRI), and positron emission tomography scans, and results are promising. Studies conducted on these imaging techniques with the DL technique for the detection of caries, osteosclerosis, root morphology, root fractures, periapical lesions, teeth identification, and other oral diseases have gained momentum recently.[5,12-19] Çok teşekkür ederim, kontrol edeceğim hemen, iyi çalışmalar dilerim. The sort of therapy to be used will depend on the number and location of missing teeth in the edentulous area. The studies on analyzing the effectiveness of artificial intelligence for tooth identification that we could locate in the literature.[2,18] We believe that our study, which assesses how well artificial intelligence performs in identifying missing teeth, will significantly advance the field of research.

A few studies have been found in the literature that have applied DL by segmenting various anatomical structures to detect missing tooth regions and for implant planning.[1] Bayrakdar *et al.*, with an AI system they developed, were able to detect canals, fossa, sinuses, and missing teeth for implant planning in CBCT images.[1,20] Liu *et al.* also evaluated the mandibular left first molar for implant planning with a proposed DL system.[21] Looking at previous studies, assessments of missing teeth have generally been made using CBCT images. However, CBCT requires a higher cost compared to panoramic radiography. In addition, these studies can only detect certain missing tooth regions and do not have fully automated methods. Therefore, it is difficult to detect more than one missing tooth region at the same time.[20,21]

Estai *et al.* aimed to detect and classify permanent teeth in patients over 18 years of age with 591 panoramic images.[2] Ninety percent of the panoramic images were used for training and the remaining 10% for validation. He reported that the resulting automated method showed high performance for tooth detection and numbering. They reported that DL will reduce the workload by assisting in the automatic filing of dental charts in general dentistry and forensics. They expressed that the study is the first step not only to detect teeth and their constituent parts, but also to detect missing teeth, dental caries, and maxillofacial problems. Therefore, they emphasized the importance of doing this first step correctly. Detection and classification of teeth will also facilitate the detection of missing teeth. The number of missing teeth in edentulous areas and their location on the crest are important in determining the type of prosthetic and orthodontic treatment, in surgical operations. Therefore, we think that rapid detection of missing teeth with AI will speed up the diagnosis process, reduce the workload and prevent possible oversight.

In a study on the detection of missing teeth using AI, Çelik *et al.* 153 panoramic radiography pre-trained a Google Net Inception v3 CNN network was preprocessed and the datasets were trained using transfer learning.[22] The success rate of the training model of the system was 94.7%, and the success rate of the images used for the test was 75%. In this study, ML performed excellently in detecting missing permanent teeth on panoramic radiographs, similar to the recently developed deep CNNs. In our study, PCA was used as a feature extractor and SVM, which is more advantageous than other classification models, was preferred.

In their study of tooth identification and missing tooth detection with 455 panoramic radiographs of Park *et al.*, which we could find in the literature, they reported that the detection of missing tooth regions could be made to a large extent.[1] While 348 of 455 images were used for training, 107 images were used to evaluate the performance of the model. As a result, the dataset is randomly split into 77.5% for training, 7.5% for validating, and 15% for testing. In our study, a larger data set with 1000 panoramic radiographs was used. 700 panoramic radiography images were used for training purposes and 300 for evaluating the performance of the model. The success of the training model was 98.14%, and the success of the images used for the test was 81.67%. The results in Table 1 and Table 2 are a testament to the models' ability to understand complex features embedded in medical images and their potential clinical utility. With the results, the accuracy of each classifier on the training and test groups was examined, shedding light on their capacity to learn from training data and generalize to new, unseen data. In summary, model selection significantly affects classification results. This research delves into the complexities of these models, illuminates the strengths and limitations of their approaches, and ultimately guides us toward an informed choice in the pursuit of excellence in medical image analysis. Depending on the specific requirements and the trade-offs between training and testing accuracy, you may want to consider factors such as model complexity, interpretability, and potential for further fine-tuning when choosing the most appropriate model for your task. Additionally, techniques such as hyperparameter tuning and data augmentation can help improve the model's performance and generalization capabilities.

In the context of the classification of images with or without missing teeth in panoramic radiographs, precision, recall, and F1-score are essential metrics used to evaluate the performance of the classification models. Precision measures the accuracy of positive predictions, indicating the ratio of correctly predicted cases of missing teeth to the total predicted as missing teeth. Recall, on the other hand, assesses the model's ability to identify all actual cases of missing teeth, determining the ratio of correctly predicted missing teeth cases to all actual missing teeth cases. F1-score is a harmonic mean of precision and recall, providing a balanced assessment of a model's overall performance, particularly useful when there is an imbalance between the classes, such as in this study with a mix of images containing and not containing missing teeth. These metrics are crucial for understanding the trade-offs between true positives and false positives and are instrumental in the comprehensive evaluation of the classification models' effectiveness. The high accuracy scores for SVM, as reported in the study, suggest a promising performance, but examining precision, recall, and F1-score can provide a more nuanced assessment of its classification capabilities.

Table 2. Classification models and classification reports

Support Vector Machine (SVM) Classifier								
	Train Classification Report				**Test Classification Report**			
	Precision	**Recall**	**F1-score**	**Support**	**Precision**	**Recall**	**F1-score**	**Support**
Deficiency	0.93	0.98	0.95	354	0.80	0.79	0.79	145
Not deficiency	0.98	0.92	0.95	344	0.80	0.81	0.81	155
Accuracy			0.95	698			0.80	300
Macro Avg	0.95	0.95	0.95	698	0.80	0.80	0.80	300
Weighted Avg	0.95	0.95	0.95	698	0.80	0.80	0.80	300
Random Forest Classifier								
Deficiency	1.00	1.00	1.00	354	0.68	0.84	0.75	145
Not deficiency	1.00	1.00	1.00	344	0.81	0.63	0.71	155
Accuracy			1.00	698			0.73	300
Macro avg	1.00	1.00	1.00	698	0.75	0.74	0.73	300
Weighted avg	1.00	1.00	1.00	698	0.75	0.73	0.73	300
Logistic Regression Classifier								
Deficiency	0.86	0.91	0.88	354	0.81	0.79	0.80	145
Not deficiency	0.90	0.84	0.87	344	0.81	0.83	0.82	155
Accuracy			0.88	698			0.81	300
Macro avg	0.88	0.88	0.88	698	0.81	0.81	0.81	300
Weighted avg	0.88	0.88	0.88	698	0.81	0.81	0.81	300
KNeighbors Classifier								
Deficiency	0.77	0.88	0.82	354	0.61	0.70	0.65	145
Not deficiency	0.85	0.74	0.79	344	0.67	0.57	0.62	155
Accuracy			0.81	698			0.64	300
Macro avg	0.81	0.81	0.81	698	0.64	0.64	0.64	300
Weighted avg	0.81	0.81	0.81	698	0.64	0.64	0.64	300
Decision Tree Classifier								
Deficiency	1.00	1.00	1.00	354	0.63	0.70	0.66	145
Not deficiency	1.00	1.00	1.00	344	0.68	0.61	0.65	155
Accuracy			1.00	698			0.65	300
Macro avg	1.00	1.00	1.00	698	0.66	0.65	0.65	300
Weighted avg	1.00	1.00	1.00	698	0.66	0.65	0.65	300
Gaussian NB Classifier								
Deficiency	0.82	0.89	0.85	354	0.62	0.92	0.74	145
Not deficiency	0.88	0.80	0.84	344	0.86	0.47	0.61	155
Accuracy			0.85	698			0.69	300
Macro avg	0.85	0.85	0.85	698	0.74	0.69	0.67	300
Weighted avg	0.85	0.85	0.85	698	0.74	0.69	0.67	300

avg: mean

CONCLUSION

ML is a subfield of artificial intelligence that can be used to detect missing teeth in panoramic radiographs. SVM is a very successful method for classifying multidimensional data. Artificial intelligence studies, which arouse great interest in medicine and dentistry, are becoming more widespread day by day, and many software and models are used to work on medical data.[23] In the field of medical image classification, the selection of the appropriate machine learning model is very important to ensure accurate and reliable diagnosis, and a lot of studies are needed to determine the best performance.

*The authors declare that there are no conflicts of interest.

REFERENCES

1. Park J, Lee J, Moon S, Lee K. Deep Learning based detection of missing tooth regions for dental implant planning in panoramic radiographic images. Appl Sci 2022; 12: 1595.
2. Estai M, Tennant M, Gebauer D, et al. Deep learning for automated detection and numbering of permanent teeth on panoramic images. DMFR 2022; 51: 20210296
3. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical ımaging. Radiographics 2017; 37: 505-515.
4. Sarker IH. Deep Learning: A comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci 2021; 2: 420.
5. Yesiltepe S, Bayrakdar IS, Orhan K, et al. A Deep learning model for ıdiopathic osteosclerosis detection on panoramic radiographs. J Med Princ Pract 2022; 31: 555-561.
6. Peldek S, Becerikli Y. Cellular automata-based binary classification. J Inf Technol 2019; 12: 43-58.
7. Rahmani AM, Yousefpoor E, Yousefpoor MS, et al. Machine learning (Ml) in medicine: Review, applications, and challenges. Mathematics 2021; 9: 2970.
8. Julian D. Designing machine learning systems with Python. Packt Publishing Ltd, 2016.
9. Raschka S, Mirjalili V. Python machine learning: Machine learning and deep learning with Python, Scikit-learn, and TensorFlow 2. Packt Publishing Ltd, 2019.
10. Peldek S, Becerikli Y. Recognition of human action in motion detected images with GMACA. J Fac Eng Archit Gazi Univ 2019; 34: 1025-1043.
11. Singh NK, Raza K. Progress in deep learning-based dental and maxillofacial image analysis: A systematic review. Expert Sys Appl 2022; 199: 116968.
12. Mohammad-Rahimi H, Motamedian SR, Rohban MH, et al. Deep learning for caries detection: A systematic review. J Dent 2022; 122: 104115.
13. Bayrakdar IS, Orhan K, Akarsu S, et al. Deep-learning approach for caries detection and segmentation on dental bitewing radiographs. Oral Radiol 2022; 38: 468-479.
14. Hiraiwa T, Ariji Y, Fukuda M, et al. A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography. DMFR 2019; 48: 20180218.
15. Hu Z, Cao D, Hu Y, et al. Diagnosis of in vivo vertical root fracture using deep learning on cone-beam CT images. BMC Oral Health 2022; 22: 382.
16. Hamdan MH, Tuzova L, Mol A, et al. The effect of a deep-learning tool on dentists' performances in detecting apical radiolucencies on periapical radiographs. DMFR 2022; 51: 20220122.
17. Sadr S, Mohammad-Rahimi H, Motamedian SR, et al. Deep Learning for Detection of Periapical Radiolucent Lesions: A Systematic Review and Meta-analysis of Diagnostic Test Accuracy. J Endod 2023; 49: 248-261.
18. Umer F, Habib S, Adnan N. Application of deep learning in teeth identification tasks on panoramic radiographs. DMFR 2022; 51: 20210504.
19. Patil S, Albogami S, Hosmani J, et al. Artificial ıntelligence in the diagnosis of oral diseases: applications and pitfalls. Diagnostics 2022; 12: 1029
20. Kurt Bayrakdar S, Orhan K, Bayrakdar IS, et al. A deep learning approach for dental implant planning in cone-beam computed tomography images. BMC Med Imaging 2021; 21: 86.
21. Liu Y, Chen Z, Chu C, Deng F. Transfer Learning via Artificial Intelligence for Guiding Implant Placement in the Posterior Mandible: An in vitro Study. Research Square 2021.
22. Çelik Ö, Odabaş A, Bayrakdar İS, Bilgir E, Akkoca F. The detection of tooth deficiency on panoramic radiography using deep learning technique: An artificial ıntelligence pilot study. Selcuk Dental Journal 2019; 6: 168-172.
23. Hasanusta M, Bayramlar OF, Palanbek S, Karan MA. Homecare services for the elderly: Turkey's perspective. Nobel Med 2021; 17: 155-165.