Data in Brief 54 (2024) 110486

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

A Dataset of apical periodontitis lesions in panoramic radiographs for deep-learning-based classification and detection

Hoang Viet Do[a], Truong Nhu Ngoc Vo[a,*], Phu Thang Nguyen[a,*], Thi Hong Lan Luong[b], Nguyen Giap Cu[c], Hoang Son Le[d]

[a] Dentistry School, Hanoi Medical University, Hanoi 010000, Vietnam
[b] Faculty of Information Technology, Hanoi University of Industry, Hanoi 010000, Vietnam
[c] Science and Technology Research and Development Centre, Thuongmai University, Hanoi 010000, Vietnam
[d] VNU Information Technology Institute, Vietnam National University, Hanoi 010000, Vietnam

ARTICLE INFO

Article history:
Received 29 February 2024
Revised 15 April 2024
Accepted 24 April 2024
Available online 5 May 2024

Dataset link: Panoramic radiographs with periapical lesions Dataset (Original data)

Keywords:
Apical periodontitis
Machine learning
Computer vision

ABSTRACT

Deep learning has been studied in recent years to identify periapical lesions- a significant indicator of periapical periodontitis in radiographs. An accurate dataset is essential for constructing an efficient learning model for detecting periapical lesions. In order to achieve this goal, we gathered and created a database of panoramic radiographs containing periapical lesions from the High-quality Dental Treatment Centre, School of Dentistry, Hanoi Medical University, between January 2016 and March 2021. Out of 16,519 radiographs, three experienced dentists identified 3,926 images of periapical lesions and annotated those lesions based on the Periapical Lesions Classification. By applying well-known data processing techniques (e.g. scaling, mirroring, and flipping), the amount of data is increased to 17,004 images through generating additional images for machine learning. The dataset has three folders: one for the original photos, one for the post-augmentation images, and the rest for the annotation of periapical lesions. The information could assist researchers in developing a predictive machine model for detecting periapical lesions in radiographs.

* Corresponding authors.
E-mail addresses: Nhungoc@hmu.edu.vn (T.N.N. Vo), Phuthang@hmu.edu.vn (P.T. Nguyen).

https://doi.org/10.1016/j.dib.2024.110486

Specifications Table

Subject	Medical Engineering, Computer Vision, Artificial Intelligence
Specific subject area	Deep-learning-based image detection and classification of periapical lesions in Panoramic Radiographs
Data format	Raw: JPG Conversion: JPG Annotation: XML
Type of data	Image
Data collection	This dataset collects images of periapical lesions in panoramic radiographs. A Panoramic X-ray device was used to capture panoramic radiograph images in the JPG image format, with a size of 2444∗1292 px and 2316×1292 px. These images were taken from January 2016 to March 2021. A total of 3926 images with periapical lesions were captured. These images were enhanced by scaling, mirroring, and noise, which expanded the dataset to 17,004 images. The folder size of the panoramic image dataset is 4.15 GB, and there are three compressed files in RAR.
Data source location	Institution: High technology of Dentistry Centre, School of dentistry, Hanoi Medical University Country: Vietnam
Data accessibility	Repository name: Mendeley Data Data identification number: DOI:10.17632/kx52tk2ddj.3 Direct URL to data: https://data.mendeley.com/datasets/kx52tk2ddj/3

1. Value of the Data

- These data are beneficial for understanding periapical lesions from panoramic radiographs with a large number of images (up to 3926 images). Dental specialists strictly check all images during each data collection and processing stage.
- Dentists can use this dataset to improve the ability to identify periapical lesions and for dental education.
- Artificial Intelligence researchers utilize this dataset for the training phase of machine learning models to classify, detect, and segment apical periodontitis lesions. This annotated dataset is used to evaluate the precision of the proposed models through the sensitivity, specificity, and other metrics.
- To the best of our knowledge, this dataset is the first attempt that defines periapical lesions to be publicly available, with a large number of 3926 original panoramic radiographs and an expanded dataset to over 17,000 images.

2. Background

Apical periodontitis is a common chronic dental disease. Statistics stated that individuals have a52 % frequency of apical periodontitis, while tooth levels have a 4–6% prevalence [1]. Apical lesions on radiographs are the most typical sign of apical periodontitis. Medical professionals may use periapical, panoramic, or Conebeam computed tomography to identify apical periodontitis. Panoramic radiographs are frequently employed to detect periapical lesions and dental disorders.

Deep learning models have been used in many studies to identify periapical lesions on panoramic radiographs [2–4]. But those studies only identified periapical lesions without classifying and determining the extent of disease damage. Furthermore, the accuracy of the models

Table 1
PAI criteria [5].

PAI Scores		Criteria
Normal (PAI ≤2)	1	Normal periapical structures
	2	Small changes in bone structures
Abnormal (PAI >3)	3	Changes in bone structure with some mineral loss
	4	Periodontitis with well-defined radiolucent areas
	5	Severe periodontitis with exacerbating features

ranges from 67.3% to 84% on a limited dataset of about 357 images to 1000 images [2–4]. In order to develop a high-quality diagnostic learning model for apical lesion detection and classification, a dataset of panoramic radiographs that is both extensive and good quality is necessary. As far as known, this dataset is the first collection of panoramic radiographs with multiple images displaying abnormal periapical lesions published worldwide.

3. Data Description

The dataset described in this article includes 3926 panoramic radiographs with different periapical lesions, and the data were enhanced by scaling, mirroring, and noise, which expanded the dataset to 17,004 images. All images were obtained with the help of dental experts. All labeled image components have been classified into three groups according to PAI classification (Table 1) [5]: PAI 3, PAI 4, and PAI5. According to previous studies, PAI 1 and PAI 2 are considered normal and eliminated [6–9]. All original images have a resolution of 235 dpi, a width of 2444 px, and a height of 1292 px.

All images were taken using the “Sirona Orthophos XG” X-ray machine- approved by the FDA in 2013 – at the High-quality Dental Treatment centre, School of Dentistry, Hanoi Medical University from January 2016 to March 2021.

The panoramic radiograph dataset can be used to diagnose the periapical lesions. For these 3926 images, we used LabelMe (Version No. 3.16.7, Massachusetts Institute of Technology, Cambridge, MA, USA) to label the periapical lesions.

Data augmentation increases the variety of the dataset and enhances the precision of deep learning algorithms [10,11]. The dataset was augmented through random scaling (0.8–1.2 times), random rotation (−90°–180°), mirroring (horizontal, vertical, horizontal vertical), and noise. After the augmented stage, the dataset obtained 17,004 photos, as indicated in Table 2.

Table 2
Description of the dataset for the panoramic images with periapical lesions.

Data enhancement	Original	Scale	Mirror	Rotation	Noise	All
Num.	3926	3926	3926	3926	1300	17,004

3.1. Component description of panoramic radiographs with periapical lesions

Fig. 1 shows the most common periapical lesions in panoramic radiographs.

3.2. Description of dataset folders

We constructed a panoramic radiograph dataset by employing 10,800 images post-data augmentation. These images were separated into folders with original photos, post-augmentation

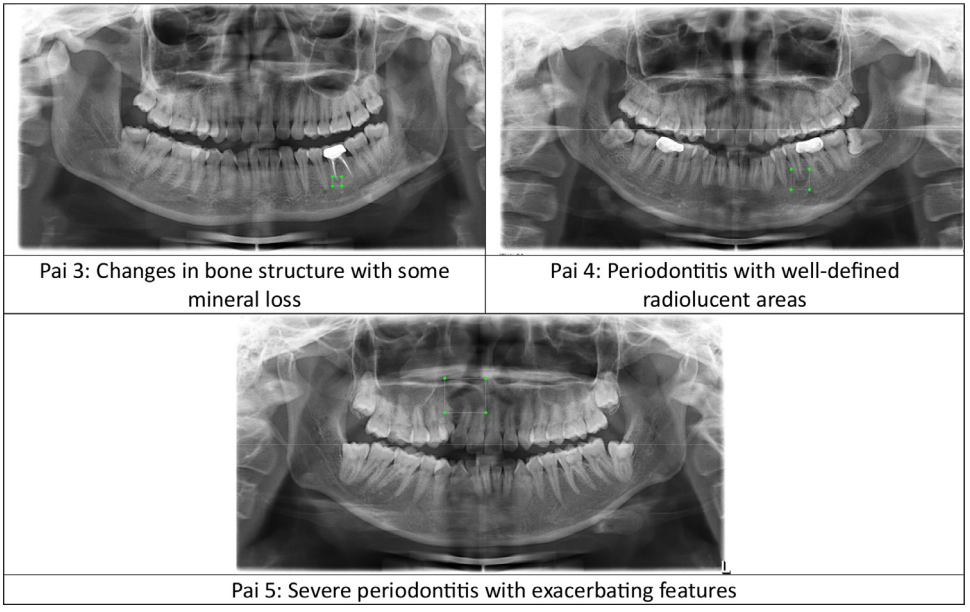

Fig. 1. The sample image of periapical lesions.

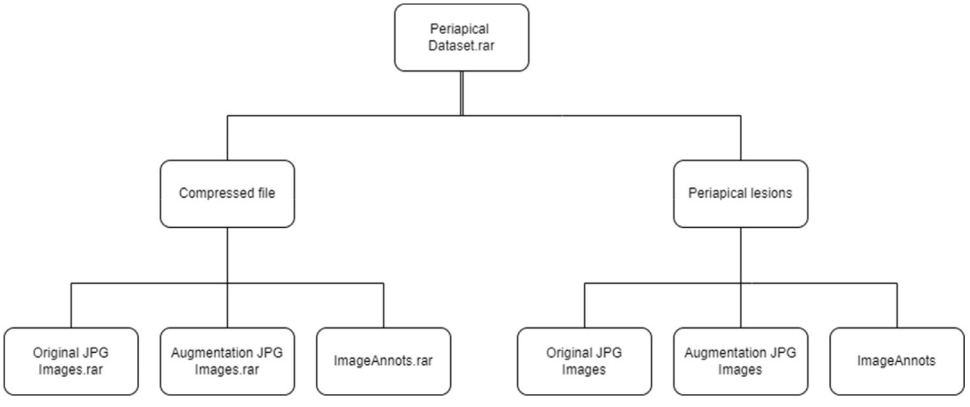

Fig. 2. Folder structure.

images, and annotation XML files (Table 3). The table shows a brief description of the dataset. The folder structure is shown in Fig. 2.

The periapical Dataset directory contains a folder named “Periapical Lesions” and a folder called “Compressed file”. The “original JPG Images” folder contains the original Images, which makes it convenient for researchers to browse and view datasets online. The “Augmentation Periapical Lesions” folder contains the images post-data augmentation. The “Compressed file” folder contains “Original JPG Images.rar”, “Augmentation JPG Images.rar,” and “ImageAnnots.rar” files that can be used to download directly. The details of the folder and file are shown in Table 4.

The Periapical lesions image dataset’s “Periapical lesions” directory comprises three main folders. Within this structure, the “Original JPG Images” folder stores the original images of panoramic radiographs with periapical lesions. The Augmentation JPG Images folder maintains

Table 3
Brief description of the dataset.

No.	Particulars	Description
1	Original Image	JPG; 2444∗1292 px; 235 dpi
2	Post augmentation image	JPG and PNG; 2444∗1292 px; 235 dpi
3	Annotation file format	XML
4	Dataset size	Size of each image: 132–869 KB Original JPG Images folder size: 817 MB Augmentation JPG Images folder size: 3.37 GB Image Annotation folder size: 11.5 MBOriginal JPG Images.rar size: 823 MB Augmentation JPG Images.rar size: 3.52 GB ImageAnnots.rar size: 6.2 MB

Table 4
Brief description of the folder/file.

No.	Name	File formats in the folder	Description
1	Original JPG Images.rar Augmentation JPG Images.rar ImageAnnots.rar	RAR	Conveniently packaged for download
2	Original JPG Images	JPG	Original Panoramic Images
3	Augmentation JPG Images	JPG	Panoramic Images Post Augmentation
4	ImageAnnot folder	XML	Labels and annotations of periapical lesions

image post-augmentation. Additionally, XML files contain the label of periapical lesions and the location of lesions in the images.

3.2.1. Original JPG images

This folder contains all 3926 images in JPG format. The folder size is 817 MB. The naming convention for images consists of five Arabic numerals, the first four digits representing the serial number of the panoramic radiographs and the last digit depicting the original image.

3.2.2. Augmentation JPG images

This folder contains all 13,078 images in JPG format. The folder size is 3.37 GB. The naming convention for images is five Arabic numerals. The first four digits represent the serial number of the original image. The last one represents the type of data enhancement, labeled from 2 to 5, where label 2 represents noise, label 3 represents random scaling, label 4 represents mirroring, and label 5 represents rotation.

For example, 00032.JPG indicates that the image was obtained by adding noise to the Panoramic image with the sequence number 3.

3.2.3. Image annotation folder

There is a folder called Annotations, which contains XML files. The name of the XML file is the same as of the JPG images. The XML file stores information about the lesion's label and its position in the JPG image regarding width, height, and depth.

In summary, in the Periapical Dataset, we have one zip file and three folders: the original images, the images post augmentation, and the XML files.

4. Experimental Design, Materials and Methods

The panoramic radiographs were taken from patients in the High-quality Dental Treatment centre, Hanoi Medical University, from January 2016 to March 2021. A total of 16,519 panoramic radiographs were collected.

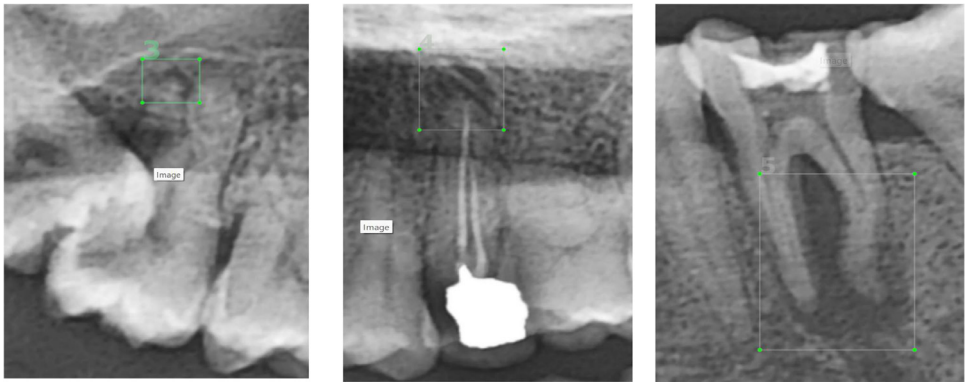

Fig. 3. The annotation box contains the whole defect.

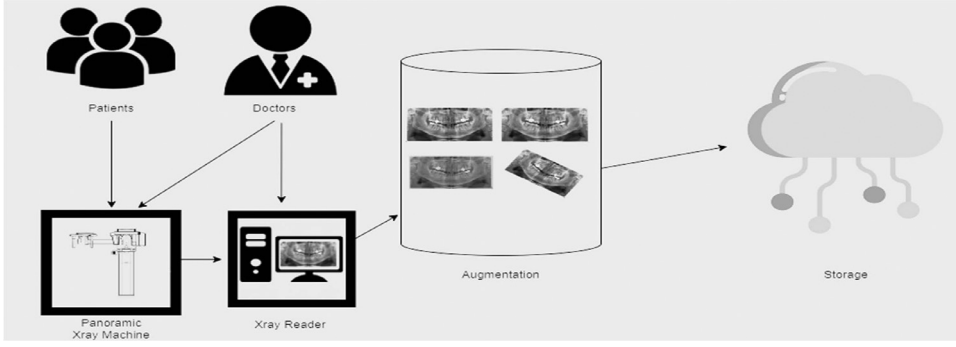

Fig. 4. Data collection and annotation workflow.

Three experienced dentists with over five years of experience assessed and classified periapical lesions in panoramic radiographs. Three dentists calibrated fifty panoramic radiographs with periapical lesions before detecting the periapical lesions in the panoramic radiograph (Figure 3). Radiographs obtained in agreement with at least two examiners will be selected with the annotation of the majority. PR with three different annotations or no lesion detected (PAI 1, 2) will be eliminated.

We used the open-source annotation software LabelMe (version 3.16.7) to annotate the panoramic radiographs dataset Three dentists mark the lesions by enclosing the entire defect with a box. The lesion areas corresponding to cases of PAI 3, PAI 4 and PAI 5 are labeled “3”, “4” and “5”, respectively. The annotations of periapical lesions were evaluated by the Intersection over Union of images (IoU) between the annotations of three dentists. If the IoU threshold is greater than 0.8 between at least two annotation boxes, we will randomly choose the annotation from one of these results. If no IoU exists between two boxes which is more than 0.8, the lesions will be eliminated. Finally, 3926 panoramic radiographs were selected for the study (Table 5). The inter-reliability of the three examiners was also considered. Data collection and annotation workflow are shown in Figure 4

4.1. Data augmentation

The panoramic images collected by dentists were then augmented by an image data enhancement program written in Python on PyCharm Community Edition. The image was randomly

Table 5
Distribution of periapical lesions according to tooth type.

	Maxillary			Mandibular			Total
	Anterior	Premolar	Molar	Anterior	Premolar	Molar	
PAI 3	271	479	817	340	580	1204	3691
PAI 4	70	330	521	109	479	308	1817
PAI 5	52	76	48	172	104	69	521
Total	393	885	1386	621	1163	1581	6029

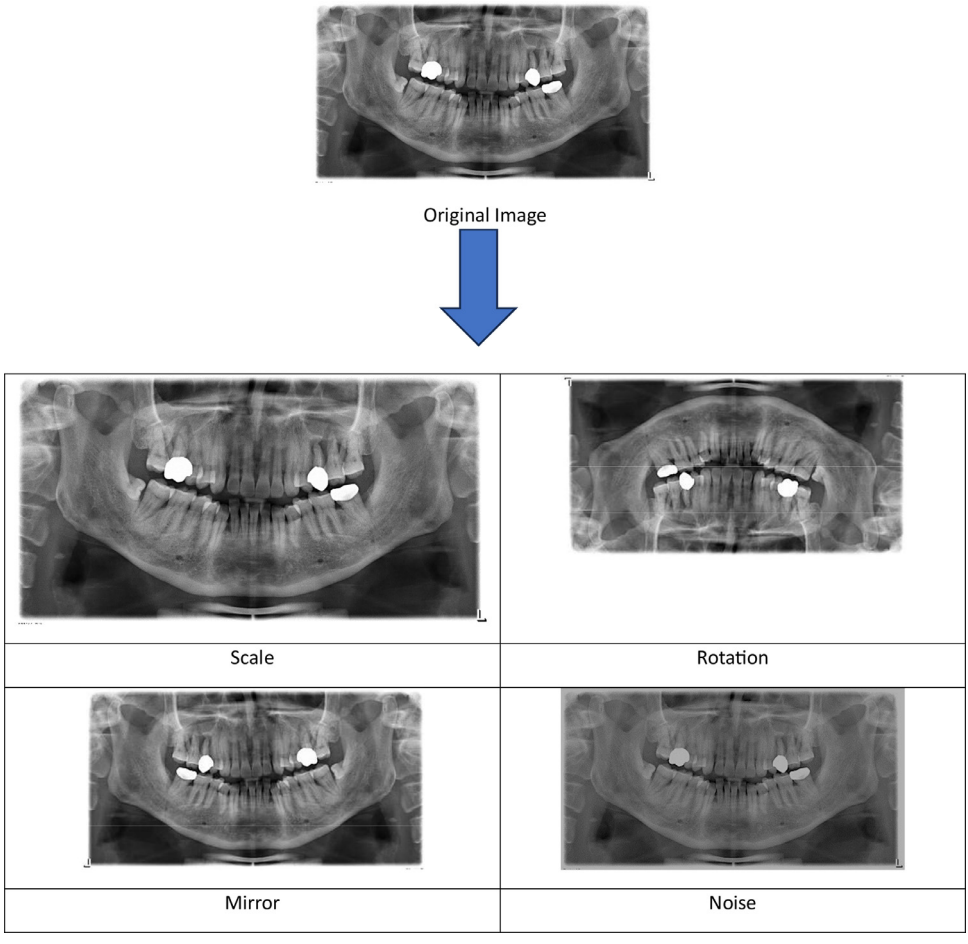

Fig. 5. Data augmentation.

scaled (0.8–1.2 times), mirrored (horizontal, vertical, horizontal vertical), rotated (−90°– 180°), and noised (Figure 5). We expanded from 6,029 apical lesions to 25,827 apical lesions (Table 6).

We saved the converted images in the “Augmentation JPG Images” folder according to naming conventions.

Table 6
Distribution of periapical lesions according to tooth type post-augmentation.

	Maxillary			Mandibular			Total
	Anterior	Premolar	Molar	Anterior	Premolar	Molar	
PAI 3	1164	2053	3505	1459	2486	5163	15,830
PAI 4	300	1416	2235	468	2054	1322	7795
PAI 5	224	326	206	734	446	296	2232
Total	1688	3795	5946	2661	4986	6781	25,857

Limitations

Not applicable.

Ethics Statement

Ethics Approval

Ethics approval was obtained from the institute research ethics committee (HMUIRB463 dated 03.08.2021).

Consent to Participate

For this type of study, formal consent is not required.

Data Availability

Panoramic radiographs with periapical lesions Dataset (Original data) (Mendeley Data).

CRediT Author Statement

Hoang Viet Do: Conceptualization, Methodology, Software, Writing – original draft, Investigation; **Truong Nhu Ngoc Vo:** Supervision, Validation, Writing – review & editing; **Phu Thang Nguyen:** Supervision, Validation, Writing – review & editing; **Thi Hong Lan Luong:** Validation, Writing – review & editing; **Nguyen Giap Cu:** Validation, Writing – review & editing; **Hoang Son Le:** Validation, Writing – review & editing.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. S. Tibúrcio-Machado, et al., The global prevalence of apical periodontitis: a systematic review and meta-analysis, Int. Endod. J. 54 (5) (2021) 712–735.
[2] B.-H. Raidan, et al., Detection of periapical lesions on panoramic radiographs using deep learning, Appl. Sci. 13 (3) (2023) 1516.
[3] B. Çelik, et al., The role of deep learning for periapical lesion detection on panoramic radiographs, Dentomaxillofac. Radiol. 52 (8) (2023) 20230118.
[4] I.-S. Song, et al., Deep learning-based apical lesion segmentation from panoramic radiographs, Imaging Sci. Dent. 52 (4) (2022) 351.
[5] D. Ørstavik, K. Kerekes, M.E. Harald, The periapical index: a scoring system for radiographic assessment of apical periodontitis, Dent. Traumatol. 2 (1) (1986) 20–34.
[6] Jurica Matijević, et al., Prevalence of apical periodontitis and quality of root canal fillings in population of Zagreb, Croatia: a crosssectional study, Croat. Med. J. 52 (6) (2011) 679–687.
[7] B. Sidaravicius, J. Aleksejuniene, H.M. Eriksen, Endodontic treatment and prevalence of apical periodontitis in an adult population of Vilnius, Lithuania, Dent. Traumatol. 15 (5) (1999) 210–215.
[8] M. Tsuneishi, et al., Radiographic evaluation of periapical status and prevalence of endodontic treatment in an adult Japanese population, Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod. 100 (5) (2005) 631–635.
[9] J. Correia-Sousa, et al., Apical periodontitis and related risk factors: Cross-sectional study, Rev. Port. Estomatol. Med. Dent. Cir. Maxilofac. 56 (4) (2015) 226–232.
[10] J. Scantlebury, et al., Data set augmentation allows deep learning-based virtual screening to better generalize to unseen target classes and highlight important binding interactions, J. Chem. Inf. Model. 60 (8) (2020) 3722–3730.
[11] A. Mikołajczyk, M. Grochowski, Data augmentation for improving deep learning in image classification problem, 2018 in2018 international interdisciplinary PhD workshop (IIPhDW)ternational interdisciplinary PhD workshop (IIPhDW), IEEE, 2018.