Clinical Oral Investigations
https://doi.org/10.1007/s00784-021-04043-y

ORIGINAL ARTICLE

Deep learning for categorization of endodontic lesion based on radiographic periapical index scoring system

Navas P. Moidu[1] · Sidhartha Sharma[1] · Amrita Chawla[1] · Vijay Kumar[1] · Ajay Logani[1]

Received: 6 February 2021 / Accepted: 21 June 2021

Abstract

Objective The study aimed to apply convolutional neural network (CNN) to score periapical lesion on an intraoral periapical radiograph (IOPAR) based on the periapical index (PAI) scoring system.

Materials and methods A total of 3000 periapical root areas (PRA) on 1950 digital IOPAR were pre-scored by three endodontists. This data was used to train the CNN model—"YOLO version 3." A total of 450 PRA was used for validation of the model. Data augmentation techniques and model optimization were applied. A total of 540 PRA on 250 digital IOPAR was used to test the performance of the CNN model.

Results A total of 303 PRA (56.11%) exhibited true prediction. PAI score 1 showed the highest true prediction (90.9%). PAI scores 2 and 5 exhibited the least true prediction (30% each). PAI scores 3 and 4 had a true prediction of 60% and 71%, respectively. When the scores were dichotomized as healthy (PAI scores 1 and 2) and diseased (PAI score 3, 4, and 5), the model achieved a true prediction of 76.6% and 92%, respectively. The model exhibited a 92.1% sensitivity/recall, 76% specificity, 86.4% positive predictive value/precision, and 86.1% negative predictive value. The accuracy, F1 score, and Matthews correlation coefficient were 86.3%, 0.89, and 0.71, respectively.

Conclusion The CNN model trained on a limited amount of IOPAR data showed potential for PAI scoring of the periapical lesion on digital IOPAR.

Clinical relevance An automated system for PAI scoring is developed that would potentially benefit clinician and researchers.

Keywords Artificial intelligence · Neural networks · Periapical diseases · Radiography · Root canal therapy

Introduction

Histological assessment, albeit not feasible in practice, is the true measure for evaluating healing of non-surgical endodontic therapy (NSET). Consequently, clinical and radiographic criteria are employed to evaluate the treatment outcome. The pre- and postoperative intraoral periapical radiographs (IOPAR) are used to assess the periapical lesion. The periapical index (PAI) is a structured and accurate scoring system comprising reference radiographs, corresponding line drawings and their associated PAI scores against which the IOPAR are scored [1] (Table 1). It is extensively utilized to evaluate the treatment outcome of NSET [2–4]. However, the process of PAI scoring is time-consuming and requires trained manpower. Further, the calculated average kappa values were only in the fair to moderate range of agreement [5].This variation in interpretation may affect the quality of the intended research since IOPAR evaluation is the basis for endodontic outcome studies [4, 6, 7].

✉ Ajay Logani
apexlogani@yahoo.co.in

Navas P. Moidu
navasktdy@gmail.com

Sidhartha Sharma
sid.aiims@gmail.com

Amrita Chawla
dr.amritachawla@gmail.com

Vijay Kumar
kumarvijay29@gmail.com

[1] Division of Conservative Dentistry and Endodontics, Centre for Dental Education and Research, All India Institute of Medical Sciences, Room No 313, Ansari Nagar, New Delhi 110029, India

Published online: 02 July 2021

Table 1 The criteria for periapical index (PAI) scoring system

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

Deep learning or convolutional neural networks (CNN) are biologically inspired neural networks which solve an equation by passing through a series of convolutional filters and simple non-linearities [8]. It has been applied in the field of medicine [9–11]. In dentistry, it has been investigated for detection of dental caries [12], periapical lesions [13–15], diagnosis of maxillary sinusitis [16], temporomandibular joint osteoarthritis [17], alveolar bone segmentation [18], and skeletal classification with lateral cephalometry [19]. CNN is a useful adjunct for diagnosis and treatment planning [20, 21].

There are no studies cited in the literature that attempted to use a deep learning model to assist in PAI scoring. In endodontics, the use of such CNN models has largely been restricted to the detection of periapical lesions [13–15]. Considering the limitations of the PAI scoring system and the effect of human bias, the purpose of this study was to use a CNN model to score the periapical lesion on an IOPAR using the PAI scoring system.

Materials and methods

Data description

This retrospective study is reported according to the checklist for artificial intelligence in medical imaging (CLAIM) guidelines [22]. Institutional review board approved the study (IECPG-215/ 28.06.2018, RT-7/30.08.2018 dated 05.09.2018). A total of 3540 periapical root areas (PRA) on 2200 digital IOPAR were obtained from subjects (n = 2200; 52% males, 48% females; mean ± SD age: 31.5 ± 10.68 year; range: 18 to 68 year) who had undergone radiological investigation for diagnostic/treatment requirement. IOPAR were exposed following the ALARA principle on a size-2 CMOS RVG sensor (Kodak RVG 5100, Eastman Kodak Company, France) held in a sensor positioning device (RINN XCP-ORA, Dentsply Int. Inc., NY, USA). These were stored and retrieved in Digital Imaging and Communications in Medicine (DICOM) format of 8 MB size from the computerized data bank. IOPAR with adequate functional quality of mature permanent mandibular teeth exhibiting healthy periapical structure or periapical lesion were included. IOPAR of mandibular third molar and teeth exhibiting coalescing periapical lesion were excluded. IOPAR were alphanumerically coded.

Training and testing of the CNN model

Expert's PAI scoring

PRA (3000) were equally divided into 120 sets. Three endodontists (VK, AC, and SS) were calibrated on a set of two IOPAR representing each class of PAI score. These IOPAR were neither repeated nor included in training/testing of the CNN model. After calibration, one set was scored on alternate days for 120 sessions. The endodontists separately viewed the IOPAR in a dimly lit room on a 22-inch monitor (1680 × 1050 pixels; Dell, TX, USA). The scores were compared, and discrepancies were settled by mutual agreement. In case of disagreement, a senior endodontist's (AL, experience greater than 25 years) score was used to train the CNN model. The IOPAR were re-scored at 1-month interval to assess the intra- and inter-observer reliability.

Annotation of the IOPAR

IOPAR were transferred to the data scientist and were hosted on a free web platform SUPERVISELY (Deep system, Yamskogo Polya, Moscow, Russia). VK, AC, and SS performed the annotation of the periapical area using the polygonal tool. Color coding was used for representing different class of PAI scores. This was followed by the final bounding box, which specified the exact location for further analysis (Fig. 1).

Training and validation

You only look once version 3 (YOLO.v3) [23] was used as the framework for the model. This enabled the combination of classic segmentation and classification task. This workflow allows the machine learning model to be less cumbersome than that of two separate models. Before feeding the annotated images into the framework, they were (pixel size 416 × 416) digitally pre-processed, so that all the images were transferred into gray scale and the pixel values of each image segment were normalized to a fixed range (0, 1). Image augmentation techniques like horizontal flipping, rotation (5° clockwise and anticlockwise), and shearing (from 0.8 to 1.2) were done during the image pre-processing stage and not on the fly to produce a total of 50,000 training data set. No vertical flipping was preferred as it could induce variability that is highly artificial and could influence the resulting model.

A total of 3000 PRA (healthy/lesion) on 1950 IOPAR were used during the training and validation phase. PAI

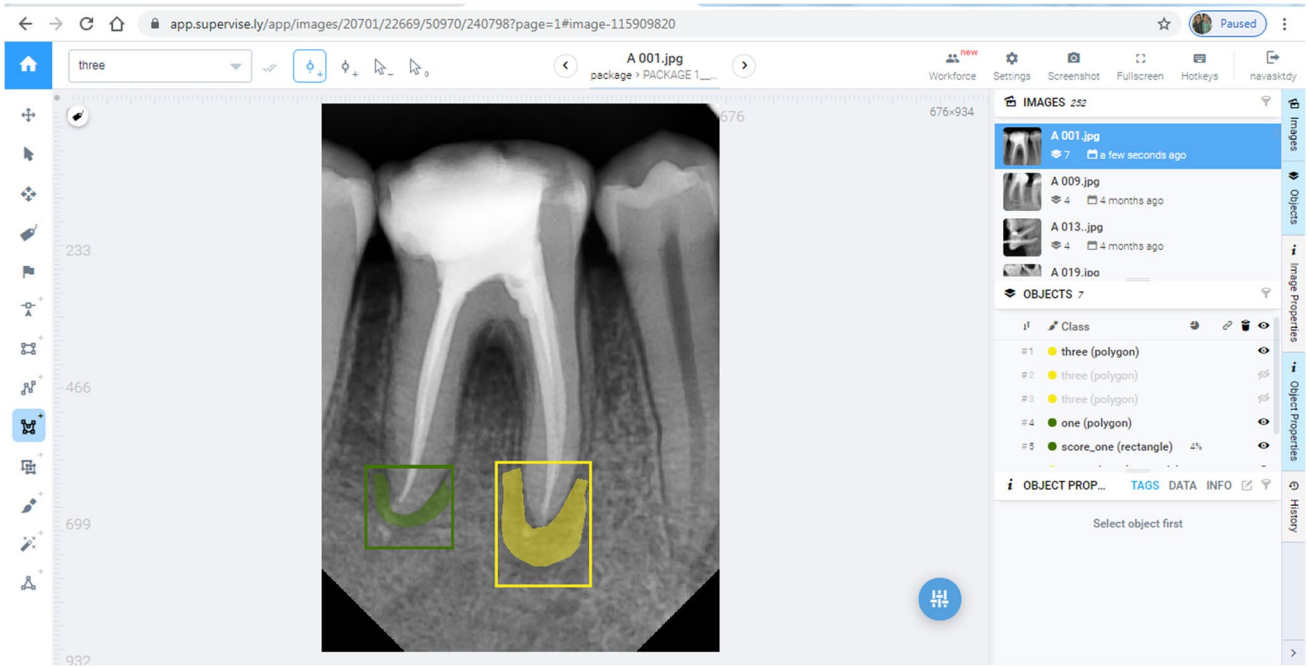

Fig. 1 Color coded annotation for different classes of PAI scores on SUPERVISELY. Areas marked in green and yellow represent the root with PAI scores 1 and 3, respectively

scores 1 to 5 were included in an equal proportion (600 PRA each) to prevent class imbalance (Table 2). To avoid overfitting and to adjust weights, 15% of the training set (450 PRA) was utilized for validation of the model. It was set up for 20,000 epochs in a batch size of 24. However, it achieved convergence at 11,000 epochs. The segmentation network utilized stochastic gradient descent with momentum optimizer with an initial learning rate at 0.001 for the first 5000 epochs and 0.01 for the remaining 6000 epochs. Sum of squared error loss function was applied for segmentation by the YOLO.v3 and cross entropy loss function was used for the classification. Total trainable parameters were 7.13 million. Detailed hyperparameter settings are provided (Table 3). The setup of the model architecture and optimization process was performed using Keras library and Python programming language. It was trained on Ubuntu 18.04 LTS operating system (Canonical Group Limited, London, UK) with Nvidia Tesla K 80 GPU graphics card (Nvidia Corporation, Santa Clara, CA).

Table 2 Data split used for the various stages of CNN model development

PAI Score	Training	Validation	Testing	Total
Score 1	510	90	110	710
Score 2	510	90	100	700
Score 3	510	90	110	710
Score 4	510	90	100	700
Score 5	510	90	120	720

Testing

A total of 540 PRA (healthy/lesion) on 250 IOPAR (Table 2) that were not included in the training or validation of the model were scored by three endodontists (VK, AC, and SS) following the previously mentioned methodology. The IOPAR were re-scored at 1-month interval to assess the intra- and inter-observer reliability. This was taken as the reference standard score. The model was evaluated for segmentation performance using Dice coefficient and mean intersection over union (Jaccard index). The former is calculated by measuring the symmetrical overlap

Table 3 Hyperparameters used for training the network

Hyperparameter	Value
Learn rate	0.001
Optimizer	SGDM
Loss	Sum-squared error
Momentum	0.9
Decay	0.0005
Batch size	24
Max. batches	500,200
Scale factor	0.1

Table 4 Segmentation accuracy by tooth type

Tooth type	Mean IoU
Mandibular incisors	0.89
Mandibular canine	0.88
Mandibular premolars	0.89
Mandibular molars	0.90

between automatedly and manually delineated regions while the latter is measured by the size of the intersection divided by the size of the union of the sample sets. In addition, AC and SS manually assessed the boundaries predicted by the model and ensured that all the bounding boxes were correctly marked. The PAI scores were further dichotomized as healthy (PAI scores 1 and 2) and diseased (PAI scores 3, 4 and 5). CNN model's predicted score and the reference standard scores were entered in an Excel sheet.

Statistical analysis

Statistical package for social sciences (Version 22.0; SPSS, Inc., Chicago, IL, USA) was used for data evaluation. Cohen kappa statistics was used to evaluate the intra- and inter-observer agreement of the endodontists. Sensitivity/recall, specificity, positive predictive value (PPV)/precision, and negative predictive value (NPV) were calculated and graded [15, 24]. The accuracy, F1 score, and Matthews correlation coefficient (MCC) were computed.

Result

The intra- and inter-observer reliability of the three endodontists was determined to be 0.76 (0.75–0.78) and 0.71 (0.70–0.73), respectively (average Cohen kappa score). The model achieved a Dice coefficient and mean intersection over union (Jaccard index) of 0.89 and 0.90, respectively. A recall and precision values of 0.89 and 0.90 were computed (Tables 4 and 5).

Across all PAI score, 303 PRA (56.11%) exhibited true prediction. PAI score 1 showed the highest true prediction (90.9%). PAI scores 2 and 5 exhibited the least true prediction (30% each). PAI scores 3 and 4 had a true prediction of 60% and 71%, respectively (Fig. 2). When the false predictions were analyzed, it was observed that PAI score 2 was interpreted as PAI score 3 (32%) and PAI score 1 (25%). Similarly, PAI score 5 was interpreted as PAI score 4 (35.8%) and PAI score 3 (20.8%) (Table 6) (Fig. 3). When the PAI scores were dichotomized as healthy (PAI scores 1 and 2) and diseased (PAI scores 3, 4, and 5), the model achieved a true prediction of 76.6% and 92%, respectively (Table 7). The CNN model achieved a 92.1% sensitivity (excellent [15, 24]), 76% specificity (good [15, 24]), 86.4% PPV (excellent [15, 24]), and 86.1% NPV (excellent [15, 24]).

The accuracy, F1 score, and Matthews correlation coefficient of the model were calculated as 86.3%, 0.89, and 0.71, respectively.

Table 5 Whole performance matrices of the CNN model

Performance matrix	Value
Recall	0.89
Precision	0.90
Dice coefficient	0.89
Mean IoU	0.90

Discussion

The assessment of the tooth's periapical status is the foundation for numerous classical studies that serve as the framework for current endodontic practice [2–4]. The PAI scoring system's established histological correlation makes it a clinically reliable index [25]. Although it appears to be straightforward, the process of PAI scoring is not without flaws. Due to the rigidly defined scoring criteria, it is susceptible to observer bias. As a result, the development of a deep learning model to automate the PAI scoring process is necessary. The intent of the present study was creation of such a machine learning model that can simultaneously identify an area of interest and classify it in accordance with a clinically useful classification system—the PAI scores. The current model, trained on a small data set, produced a range of true predictions (30–90%) across the five classes of PAI score.

The model exhibited the highest true prediction (90.9%) for PAI score 1. Hence it is well trained to interpret healthy periapical tissue. The true prediction for PAI score 2 was 30%. Thirty-five percent of the total error (25/70) in this class was interpreted as PAI score 1. Although theoretically, the distinction between PAI scores 1 and 2 is desirable, outcome studies [3, 26] have dichotomized these scores as "healthy." Applying a similar strategy, i.e., when PAI scores 1 and 2 were dichotomized, the model achieved a true prediction of 76.6%. Hence, when the model would be applied to a clinical situation, the interpretation regarding the outcome would not be impacted. The CNN model misinterpreted PAI score 2 as 3 (45% of total error). This was an area of concern because score 2 is considered as "healthy" and a score 3 is considered "diseased." This error could be because PAI score 2 is represented as "small changes in bone structure" and PAI score 3 is "change in bone structure with some mineral loss" [1]. This difference is so infinitesimal that even an experienced endodontist can erroneously interpret them.

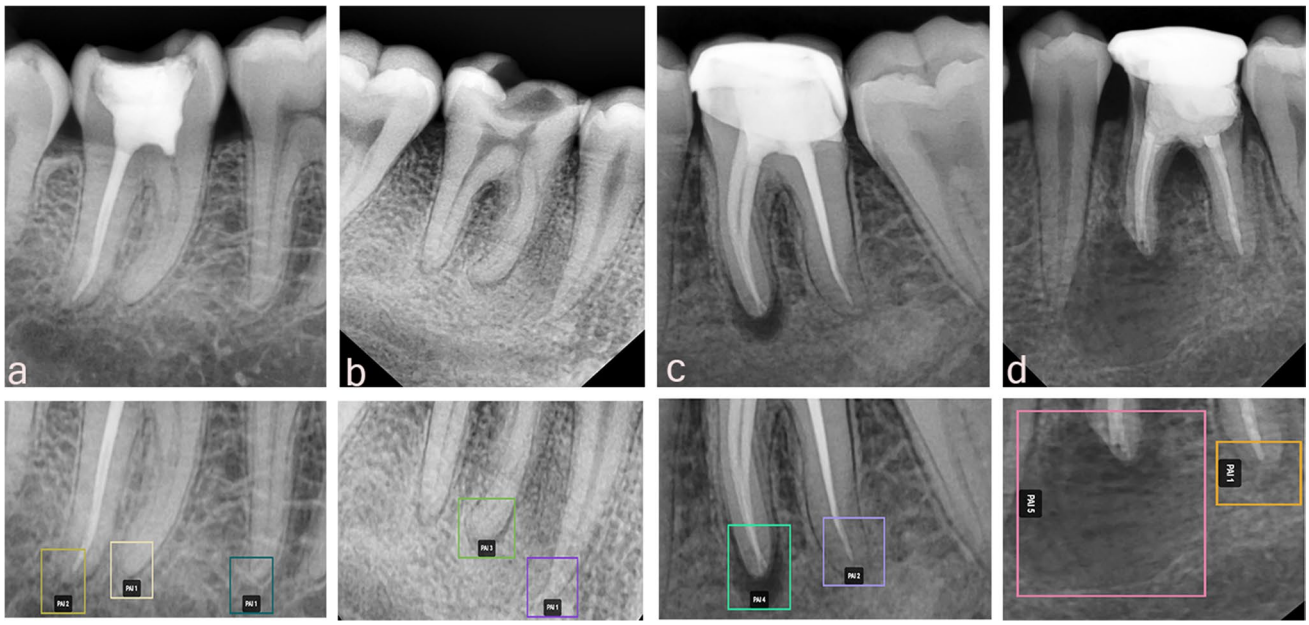

Fig. 2 CNN model exhibiting true prediction across all classes of PAI scores. **a** PAI score 1 (mesial root of second molar and distal root of first molar) and PAI score 2 (distal root of second molar). **b** PAI score 3 (mesial root of first molar) and PAI score 1 (second premolar). **c** PAI score 4 (mesial root of first molar) and PAI score 2 (distal root of first molar). **d** PAI score 5 (mesial root of first molar) and PAI score 1 (distal root of first molar)

To overcome this, the CNN model would need to be trained with an increased data set pertaining to this class of PAI score. On the contrary, the ability of the CNN model to predict PAI score 3 was acceptable as only 6.8% and 9% of the PRA (3/44 and 4/44) were interpreted as PAI scores 1 and 2, respectively. The prediction for PAI score 4 was reliable (71% true prediction). The model exhibited a limited ability to distinguish between apical periodontitis with exacerbating features (PAI score 5) from a well-defined radiolucent lesion (PAI score 4). This was another shortcoming of the present model. However, the clinical significance of this discrepancy may not be relevant as both scores are considered as "diseased." When PAI scores 3, 4, and 5 were dichotomized as diseased, the model achieved a true prediction of 92%. When the false predictions were analyzed across the classes of PAI scores, the model had the tendency to overscore by one class; for example, PAI score 2 was interpreted as score 3 and PAI score 3 as score 4 (Table 3). Even though these were false predictions, they were similar to the interpretation directions of the PAI scoring system wherein when in doubt, the clinician is always advised to assign the higher score [1]. For some unknown reason, the developed model also followed this pattern, and the borderline cases were automatically given the next higher score. This pattern was reversed for PAI score 5, where the maximum erroneous interpretations were as PAI score 4. The model achieved a segmentation performance of close to 1 through Dice coefficient and intersection over union. A value of 1 signifies a perfect overlap of the CNN prediction and the expert annotation [27]. Converting the multi-class to binary classification,

Table 6 Performance of the CNN model to predict PAI scores

PAI scores	Prediction status						
	True prediction Number/percentage	False prediction Number/percentage (PAI scores)					Total
		1	2	3	4	5	
Score 1	100/(90.9)	0	6[#] /(5.5)	4/(3.6)	0	0	110
Score 2	30/(30)	25/(25)	0	32[#]/(32)	13/(13)	0	100
Score 3	66/(60)	3/(2.7)	4/(3.6)	0	32[#]/(29)	5/(4.54)	110
Score 4	71/(71)	3/(3)	NIL	6 /(6)	0	20[#]/(20)	100
Score 5	36/(30)	3/(2.5)	13/(10.8)	25/(20.8)	43[#]/(35.8)	0	120

[#]Maximum error in classification

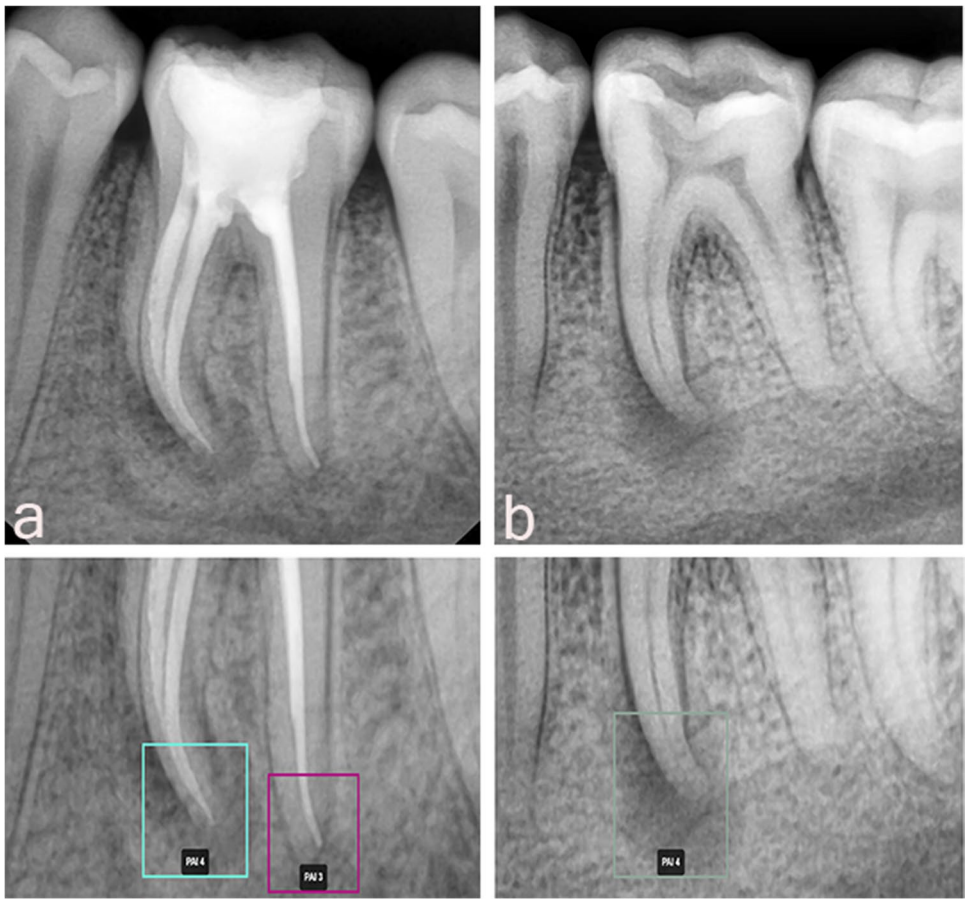

Fig. 3 CNN model exhibiting false predictions. **a** PAI 2 interpreted as PAI 3 (distal root of first molar). **b** PAI 5 interpreted as PAI 4 (mesial root of first molar)

the model achieved an accuracy of 86.29%, F-1 score of 0.89, and MCC of 0.71. MCC score of 1 signifies a perfect model [28]. The ability of the model to detect diseased PRA (sensitivity) was 91%. However, the ability of the model to detect disease-free PRA (specificity) was 77%. This was because PAI score 2 (healthy) was falsely predicted as PAI score 3(diseased) in 32% of the total PRA assessed.

This CNN model was trained to perform multiple classifications by providing detailed annotations of all possible subclasses of the periapical lesion. This prevented oversimplification and the within class heterogeneity [29]. Further by including comparable data set, the class imbalance was prevented [30]. This method provided the required flexibility to convert the model into a two class–based system in the future, if the need arises. Based on the experience of this model, it can be suggested that such an automated system for PAI scoring has the potential for use in clinical decision-making and nullify human bias. Additionally, the requirement of trained manpower and the time taken to score the IOPAR can be reduced substantially. This would be specifically beneficial for large-scale epidemiological or retrospective radiological studies, where the interpretation of voluminous data by the human observer could be a daunting task.

The shortcoming of the present study was primarily the limited training data set, a fact which has been highlighted in the literature for similar studies [31]. In the present study, the data set was limited due to the strict inclusion criteria.

Table 7 Performance of the CNN to predict a healthy periapical root area (PAI scores 1 and 2) and diseased periapical root area (PAI scores 3, 4, and 5)

Category	Prediction status						
	True prediction Number/percentage	False prediction Number/percentage (PAI scores)					Total
		1	2	3	4	5	
Healthy	161 /(76.6)	0	0	36[#]/(17.1)	13/(6.1)	0	210
Diseased	304 /(92)	9/(2.7)	17[#]/(5.3)	0	0	0	330

[#]Maximum error in classification

Mandibular teeth were only included in the study due to the possible superimposition of the anatomical structures in the maxillary region that could have hindered the scoring of the periapical lesion. Since the CNN model was being trained for the first time for such a purpose, it was decided to minimize this confounding factor. Further, the data was limited as it was retrieved from a single center. A multicenter data collection would diversify and increase the number of training data. This model was trained to interpret only digital IOPAR. Its performance on scanned conventional IOPAR would need to be assessed. The parameters of classification task (sensitivity, specificity, PPV, NPV, accuracy, and MCC) were evaluated only at the image level which was in accordance with a previous study (17). However, evaluation at pixel/lesion level would have highlighted the detection accuracy of each class [32, 33]. This may be considered as another shortcoming of the present study. The periapical status of the tooth is one of the inputs required for complex decision-making in clinical endodontic practice. Integrating non-imagery data like history, clinical sign, and symptoms with the imagery data in the deep learning system could be attempted in the future to improve diagnostics, decision-making, and outcome.

In the present study, a CNN model was successfully trained and applied to score the periapical lesion of mandibular teeth on IOPAR based on the PAI scoring system. The model exhibited excellent sensitivity, positive predictive value, and negative predictive value. A good specificity was achieved.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00784-021-04043-y.

Acknowledgements The authors would like to thank the contributions of Dr Amandeep Kaur, Dr Aakriti Saini, and Dr Tabiyar Krunal; data scientists Ankan Dutta and Praveen Vijai; and Assistant Professors Ameer PM and Deepak S.

Declarations

Ethics approval Ethics approval is obtained from Institute research ethics committee (IECPG-215/28.06.2018, RT-7/30.08.2018 dated 05.09.2018). This article does not contain human participants or animals performed by any of the authors.

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

Conflict of interest The authors declare no competing interests.

References

1. Ørstavik D, Kerekes K, Eriksen HM (1986) The periapical index: a scoring system for radiographic assessment of apical periodontitis. Dent Traumatol 2:20–34
2. Eriksen HM, Bjertness E (1991) Prevalence of apical periodontitis and results of endodontic treatment in middle-aged adults in Norway. Dent Traumatol 7:1–4
3. Gumru B, Tarcin B, Pekiner FN, Ozbayrak S (2011) Retrospective radiological assessment of root canal treatment in young permanent dentition in a Turkish subpopulation: root fillings in young permanent teeth. Int Endod J 44:850–856
4. Kirkevang LL, Ørstavik D, Bahrami G, Wenzel A, Vaeth M (2017) Prediction of periapical status and tooth extraction. Int Endod J 50:5–14
5. Tarcin B, Gumru B, Iriboz E, Turkaydin DE, Ovecoglu HS (2015) Radiologic assessment of periapical health: comparison of 3 different index systems. J Endod 41:1834–1838
6. Friedman S, Abitbol S, Lawrence HP (2003) Treatment outcome in endodontics: the Toronto study. Phase 1: initial treatment. J Endod 29:7
7. Saini HR, Tewari S, Sangwan P, Duhan J, Gupta A (2012) Effect of different apical preparation sizes on outcome of primary endodontic treatment: a randomized controlled trial. J Endod 38:1309–1315
8. LeCun Y, Boser BE, Denker JS, Henderson D, Howard RE, Hubbard WE, et al. (2019) Handwritten Digit Recognition with a Back-Propagation Network. :9
9. Gabr RE, Coronado I, Robinson M, Sujit SJ, Datta S, Sun X et al (2019) Brain and lesion segmentation in multiple sclerosis using fully convolutional neural networks: a large-scale study. MultScler 135245851985684
10. Zhu N, Najafi M, Han B, Hancock S, Hristov D (2019) Feasibility of image registration for ultrasound-guided prostate radiotherapy based on similarity measurement by a convolutional neural network. Technol Cancer Res Treat 18:153303381882196
11 Ozawa T, Ishihara S, Fujishiro M, Kumagai Y, Shichijo S, Tada T (2020) Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks. Therap Adv Gastroenterol 13:175628482091065
12. Cantu AG, Gehrung S, Krois J, Chaurasia A, Rossi JG, Gaudin R et al (2020) Detecting caries lesions of different radiographic extension on bitewings using deep learning. J Dent 100:103425
13. Ekert T, Krois J, Meinhold L, Elhennawy K, Emara R, Golla T et al (2019) Deep learning for the radiographic detection of apical lesions. J Endod 45:917–22.e5
14. Orhan K, Bayrakdar IS, Ezhov M, Kravtsov A, Özyürek T (2020) Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. Int Endod J 53:680–689
15. Setzer FC, Shi KJ, Zhang Z, Yan H, Yoon H, Mupparapu M et al (2012) Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam computed tomographic images. J Endod 46:987–993
16. Kim H-G, Lee KM, Kim EJ, Lee JS (2019) Improvement diagnostic accuracy of sinusitis recognition in paranasal sinus X-ray using multiple deep learning models. Quant Imaging Med Surg 9:942–951
17. Lee KS, Kwak HJ, Oh JM, Jha N, Kim YJ, Kim W, Baik UB, Ryu JJ (2020) Automated detection of TMJ osteoarthritis based on artificial intelligence. J Dent Res 99(12):1363–1367. https://doi.org/10.1177/0022034520936950
18. Nguyen KCT, Duong DQ, Almeida FT, Major PW, Kaipatur NR, Pham TT et al (2020) Alveolar bone segmentation in intraoral ultrasonographs with machine learning. J Dent Res 99:1054–1061
19. Yu HJ, Cho SR, Kim MJ, Kim WH, Kim JW, Choi J (2020) Automated skeletal classification with lateral cephalometry based on artificial intelligence. J Dent Res 99:249–256
20. Jung SK, Kim TW (2016) New approach for the diagnosis of extractions with neural network machine learning. Am J Orthod Dentofacial Orthop 149:127–133

21. Saghiri MA, Garcia-Godoy F, Gutmann JL, Lotfi M, Asgar K (2012) The reliability of artificial neural network in locating minor apical foramen: a cadaver study. J Endod 38:1130–1134
22. Mongan J, Moy L, Kahn CE (2020) Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. RadiolArtifIntell 2:e200029
23. Redmon J, Farhadi A. YOLOv3: an incremental improvement. arXiv2018;180402767 [cs]. Available at: http://arxiv.org/abs/1804.02767. Accessed 2 Aug 2020
24. Leonardi Dutra K, Haas L, Porporatti AL, Flores-Mir C, Nascimento Santos J, Mezzomo LA et al (2016) Diagnostic accuracy of cone-beam computed tomography and conventional radiography on apical periodontitis: a systematic review and meta-analysis. J Endod 42:356–364
25. Brynolf I (1967) Histological and roentgenological study of the periapical region of human upper incisors. Odontol Revy 18:1–176
26. Ørstavik D, Qvist V, Stoltze K (2004) A multivariate analysis of the outcome of endodontic treatment. Eur J Oral Sci 112:224–230
27. Zou KH, Warfield SK, Bharatha A, Tempany CMC, Kaus MR, Haker SJ et al (2004) Statistical validation of image segmentation quality based on a spatial overlap index 1. Acad Radiol 11:178–89
28. Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom 21:6
29. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M et al (2017) A survey on deep learning in medical image analysis. Med Image Anal 42:60–88
30. Haibo H, Garcia EA (2009) Learning from imbalanced data. IEEE Trans Knowl Data Eng 21:1263–1284
31 Schwendicke F, Golla T, Dreher M, Krois J (2019) Convolutional neural networks for dental image diagnostics: a scoping review. J Dent 91:103226
32. Ariji Y, Yanashita Y, Kutsuna S, Muramatsu C, Fukuda M, Kise Y et al (2019) Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. Oral Surg Oral Med Oral Pathol Oral Radiol 128:424–430
33. Muramatsu C, Morishita T, Takahashi R, Hayashi T, Nishiyama W, Ariji Y et al (2020) Tooth detection and classification on panoramic radiographs for automatic dental chart fling: improved classification by multi-sized input data. Oral Radiol. https://doi.org/10.1007/s11282-019-00418-w. Online ahead of print