

INPLASY

Diagnostic Performance and Interpretability of Convolutional Neural Networks Using Grad-CAM in Orthodontic Diagnosis: A Systematic Review

INPLASY2026100004

doi: 10.37766/inplasy2026.10.0004

Received: 1 October 2026

Published: 1 October 2026

Noriega, A; Casillas, M; Ventura, E; Luna, A; Arana, Y; Perez, S.

Corresponding author:

Ana Noriega

draananoriega@gmail.com

Author Affiliation:

Benemerita Universidad Autonoma de Puebla.

ADMINISTRATIVE INFORMATION**Support** - No support.**Review Stage at time of this submission** - Piloting of the study selection process.**Conflicts of interest** - None declared.**INPLASY registration number:** INPLASY2026100004**Amendments** - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 1 October 2026 and was last updated on 1 October 2026.**INTRODUCTION**

Review question / Objective ^{**}(P) Participants:^{**} Subjects evaluated using imaging studies for orthodontic diagnostic purposes.

^{**}(I) Intervention types:^{**} Application of CNNs supplemented by Grad-CAM for orthodontic diagnosis and the visual identification of regions of interest.

^{**}(C) Control intervention:^{**} Reference diagnosis established via clinical criteria, imaging findings, or specialist assessment, where available.

^{**}(O) Outcome:^{**} Diagnostic performance of the CNNs, determined by the accuracy reported in the studies, as well as the interpretability of their results via activation maps generated using Grad-CAM.

Rationale The increasing adoption of CNNs in orthodontic diagnosis has generated evidence

regarding their ability to analyze various imaging modalities; however, adequate diagnostic performance alone does not reveal which anatomical features or regions influence the model's decision. In this context, Grad-CAM serves as an interpretability tool that allows for the visualization of the regions of interest considered by the CNN during the classification process. Therefore, it is necessary to systematically synthesize the available evidence on the diagnostic accuracy of these models and the interpretability provided by Grad-CAM, in order to identify the results achieved, the relevant anatomical regions, and the existing limitations regarding their application in orthodontic diagnosis.

Condition being studied Dentofacial and craniofacial conditions relevant to orthodontic diagnosis and assessed through diagnostic imaging. The review focuses on studies applying convolutional neural networks (CNNs) for diagnostic classification and Grad-CAM for visualization and interpretation of the anatomical regions contributing to model predictions.

METHODS

Search strategy The main keywords used were: “Convolutional Neural Network,” “CNN,” “CNNs,” “Deep Learning,” “Grad-CAM,” “Gradient-weighted Class Activation Mapping,” “orthodontic,” “orthodontics,” “CBCT,” “panoramic radiograph,” and “lateral cephalogram.” It should be noted that the term “diagnosis” was not included as a keyword, as its incorporation into the Web of Science search strategy yielded a significantly higher number of results, including studies that were not directly related to the objective of this review.

Participant or population Subjects evaluated using imaging studies for orthodontic diagnostic purposes.

Intervention Application of CNNs supplemented by Grad-CAM for orthodontic diagnosis and the visual identification of regions of interest.

Comparator Reference diagnosis established via clinical criteria, imaging findings, or specialist assessment, where available.

Study designs to be included Original studies evaluating the diagnostic performance of convolutional neural networks (CNNs) combined with Grad-CAM in orthodontic diagnostic imaging will be included. Eligible designs will comprise retrospective and prospective observational studies, as well as diagnostic model development and/or validation studies. Only studies reporting diagnostic accuracy and using Grad-CAM for model interpretability will be considered.

Eligibility criteria ### Inclusion criteria

- Articles with the primary objective of evaluating the combined application of CNNs and Grad-CAM in orthodontic diagnosis.
- Studies using images applicable to orthodontic diagnosis.
- Studies reporting accuracy as a measure of CNN diagnostic performance.
- Publications available in full text and written in English.

Exclusion criteria

- Review articles, systematic reviews, or meta-analyses.
- Studies that do not use CNNs and Grad-CAM in combination.
- Studies unrelated to orthodontic diagnosis.

- Studies focused exclusively on image segmentation or cephalometric landmark detection.
- Studies that do not report accuracy as a measure of CNN diagnostic performance.
- Publications not available in full text or written in languages other than English.

Information sources A systematic search was conducted for studies published between 2021 and 2026, focusing on English-language articles indexed in electronic databases such as PubMed, Web of Science, Springer, and Scopus. The search aimed to identify studies focusing on the application of CNNs integrated with Grad-CAM as diagnostic support tools in orthodontics.

Main outcome(s) The primary outcomes of this systematic review will be the diagnostic performance and interpretability of convolutional neural networks (CNNs) combined with Grad-CAM in orthodontic diagnosis. Diagnostic performance will be assessed primarily through accuracy, reported as a percentage or proportion of correctly classified cases. Only studies reporting accuracy will be eligible for inclusion. Additional diagnostic performance measures, such as sensitivity, specificity, precision, F1-score, and area under the receiver operating characteristic curve (AUC), will be extracted when available.

Interpretability will be assessed through the Grad-CAM activation maps reported by the included studies, focusing on the anatomical regions highlighted as relevant to the CNN predictions and their relationship with the diagnostic task.

No specific follow-up period or assessment time point will be required, as the review focuses on the diagnostic performance of CNN models rather than longitudinal clinical outcomes. Performance measures will be extracted from the test or independent validation datasets reported in each study, when available.

Quality assessment / Risk of bias analysis The methodological quality and risk of bias of the included studies will be assessed using the Quality Assessment of Diagnostic Accuracy Studies 3 (QUADAS-3) tool. The assessment will focus on risk of bias and concerns regarding the applicability of the diagnostic accuracy estimates reported in the included studies.

Given the artificial intelligence-based nature of the review, particular attention will also be paid to methodological aspects related to CNN development and evaluation, including dataset

characteristics, separation of training, validation, and test sets, potential data leakage, reference standards, and internal or external validation. The assessment will be performed independently by two reviewers, and disagreements will be resolved through discussion and consensus.

Strategy of data synthesis A narrative and descriptive synthesis of the included studies will be performed. Study characteristics will be summarized in tables, including study design, sample size, imaging modality, orthodontic diagnostic task, CNN architecture, dataset characteristics, and validation approach.

Diagnostic performance will be synthesized based on the accuracy reported for each CNN model, preferably using results obtained from test or independent validation datasets. Interpretability findings will be synthesized descriptively according to the anatomical regions highlighted by Grad-CAM activation maps and their relevance to the corresponding orthodontic diagnostic task.

Due to the expected heterogeneity among studies regarding diagnostic tasks, imaging modalities, CNN architectures, datasets, and methodological approaches, a quantitative meta-analysis is not initially planned.

Subgroup analysis If sufficient comparable studies are available, subgroup analyses will be considered according to imaging modality and orthodontic diagnostic task. Within each subgroup, the reported accuracy of CNN models and the anatomical regions highlighted by Grad-CAM will be descriptively compared. Subgroup analyses will only be performed when an adequate number of methodologically and clinically comparable studies is available.

Sensitivity analysis No formal sensitivity analysis is planned, as the review will primarily use a narrative and descriptive synthesis and no quantitative meta-analysis is initially intended. The methodological quality and risk of bias of the included studies will be considered when interpreting the findings.

Language restriction English.

Country(ies) involved Mexico.

Keywords “Convolutional Neural Network,” “CNN,” “CNNs,” “Deep Learning,” “Grad-CAM,” “Gradient-weighted Class Activation Mapping,” “orthodontic,” “orthodontics,” “CBCT,” “panoramic radiograph,” and “lateral.

Contributions of each author

Author 1 - Ana Noriega - Drafted the manuscript and registered the systematic review.

Email: draananoriega@gmail.com

Author 2 - Miguel Casillas - Provided the main idea for the systematic review, proposed registering the review in Inplasy, made modifications for PICO.

Author 3 - Elias Ventura - Suggested to focus the review on Artificial Intelligence as CNNs and GRAD-CAM.

Author 4 - Alma Luna - Contributed with the recommendation of applying the review in orthodontic diagnosis.

Author 5 - Yoaly Arana - Read and provided feedback.

Author 6 - Sabrina Perez - Approved the final manuscript.