



Reproducibility of open-source software for dental artificial intelligence

Reproducibilidad del software de código abierto para la inteligencia artificial odontológica

Flor M Escalante Freitez¹  
Roberto Antonio Vega Bonilla²
Diana Belkis Moreno Benítez²

¹ Universidad de Carabobo, Venezuela

² Universidad Nacional de Chimborazo, Facultad de Ciencias de la Salud. Riobamba, Ecuador.

Submitted: 03-05-2026 | Revised: 30-06-2026 | Accepted: 25-08-2026 | Published: 27-08-2026

Corresponding Author: Flor M Escalante Freitez 

ABSTRACT

Introduction: Open-source software underpins the reproducibility of dental artificial intelligence, yet the state of its licensing and documentation is unclear. **Methods:** We retrieved public GitHub repositories through the GitHub search API using dental artificial intelligence queries, extracted repository metadata, and inspected the root file listing of the most popular repositories. **Results:** Of 78 repositories, 23 (29.5%) carried a license and 74 (94.9%) a description; among the 30 most popular, 30 (100%) had a README and 8 (26.7%) a dependency specification; the median star count was 6 and none were archived. **Conclusions:** Most open-source dental artificial intelligence software lacks a license and a dependency specification, which limits reuse and reproduction despite near-universal README documentation.

Keywords: artificial intelligence; dental informatics; open-source software; reproducibility; machine learning.

RESUMEN

Introducción: El software de código abierto sustenta la reproducibilidad de la inteligencia artificial odontológica, pero el estado de sus licencias y documentación no está claro. **Métodos:** Se recuperaron repositorios públicos de GitHub mediante la API de búsqueda de GitHub con consultas de inteligencia artificial odontológica, se extrajeron los metadatos de los repositorios y se inspeccionó la lista de archivos raíz de los repositorios más populares. **Resultados:** De 78 repositorios, 23 (29,5 %) tenían licencia y 74 (94,9 %) descripción; entre los 30 más populares, 30 (100 %) tenían README y 8 (26,7 %) un archivo de dependencias; la mediana de estrellas fue 6 y ninguno estaba archivado. **Conclusiones:** La mayor parte del software de código abierto de inteligencia artificial odontológica carece de licencia y de un archivo de dependencias, lo que limita la reutilización y la reproducción pese a una documentación README casi universal.

Palabras clave: inteligencia artificial; informática odontológica; software de código abierto; reproducibilidad; aprendizaje automático.

INTRODUCTION

Artificial intelligence has moved from proof-of-concept demonstrations to clinical dental practice, where convolutional networks now segment panoramic radiographs, detect carious lesions, and classify periapical and periodontal findings. The pace of this transition has been rapid, and the number of published models, datasets, and software tools for dental tasks has grown steadily over the past several years, so that a clinician or researcher seeking an automated solution now faces an abundance of choices rather than a scarcity. Schwendicke, Samek and Krois (1) described the chances and challenges of these systems for dentistry, noting that the technology promises to standardize diagnosis while raising questions about generalization and accountability.

Katsumata (2) reviewed the diagnostic imaging applications in which deep learning performs closest to clinician-level accuracy, and Singh and Raza (3) mapped the progress of dental and maxillofacial image analysis from early classification tasks to segmentation and detection. Sivari et al.(4) and AbuSalim et al.(5) reached similar conclusions in their systematic reviews of dental anomaly diagnosis and dental informatics, respectively, while Abd Rahim et al.(6) measured the evidence maturity and translational readiness of the field through a bibliometric lens.

Across this literature, however, the emphasis falls almost entirely on model performance; the software artifacts that would let another team rebuild, verify, or reuse the model receive far less attention. This asymmetry is especially consequential in a clinical field, where the software that reaches the chairside must behave as its published evaluation claims, and where trust cannot be established without the possibility of independent inspection and reuse.

The broader machine-learning community has documented a reproducibility problem that extends to health care. Hutson (7) reported that artificial intelligence faces a reproducibility crisis, and Hutson (8) asked whether the field had become alchemy because published models could not be reconstructed from their descriptions alone.

Gibney (9) argued that machine learning could fuel a reproducibility crisis in science, and Kapoor and Narayanan (10) demonstrated that data leakage, rather than genuine signal, explains a substantial share of reported performance in machine-learning-based science. Within medicine specifically, Beam, Manrai and Ghassemi (11) identified the challenges to reproducing machine-learning models in health care, and McDermott et al.(12) found that only a minority of published health models reported the information needed for independent verification. These authors established that a reported result is not evidence until the code, data, and environment that produced it can be inspected, and that a field that publishes accuracy figures without those artifacts is accumulating claims rather than building a trustworthy base of knowledge.

Recognition of these problems has produced standards and checklists that now guide the field. Heil et al.(13) proposed reproducibility standards for machine learning in the life sciences, and Wilkinson et al.(14) set out the FAIR guiding principles so that data and software are findable, accessible, interoperable, and reusable. In dentistry, Schwendicke et al.(15) published a checklist for authors, reviewers, and readers of artificial-intelligence papers, and Rokhshad et al.(16) added an ethical framework and checklist for the same purpose.

Brinz et al.(17) examined the legal and privacy-preserving dimensions of data sharing for responsible dental artificial intelligence. These instruments, however, presuppose that source code is actually available and adequately described; they do not tell us whether the growing body of open-source dental artificial-intelligence software is itself licensed, documented, and reproducible in practice.

Open science and open-source software have been proposed as remedies for these weaknesses, and the free hosting platform GitHub has become the de facto venue where such software is shared and discovered. Prior empirical work outside dentistry suggests that the promise of open source is only partially realized in practice. Jebari et al.(18) examined containerization practices in open-source machine-learning projects and found that reproducible execution environments are frequently absent, and Jansen et al.(19) reported the same for computational life sciences and proposed container-based solutions to close the gap. These studies measure software in adjacent fields, but none has asked the analogous question for the software that now underpins dental artificial-intelligence applications, and it is this question that the present study addresses directly.

Open-source software carries particular weight in dental artificial intelligence because proprietary clinical tools are typically distributed as closed systems whose internal models cannot be examined, whereas open repositories at least expose the code that produced a result. The value of openness, however, is realized only when the exposed code is accompanied by the license that permits its use, the description that explains its purpose, and the dependency specification that allows it to run. A public repository that omits these elements offers the appearance of transparency

without its practical benefit, and a field that increasingly relies on such software therefore needs an empirical account of how often these elements are actually present.

What Schwendicke, Samek and Krois,⁽¹⁾ Heil et al.,⁽¹³⁾ and McDermott et al.⁽¹²⁾ did not answer is the empirical state of the open-source dental artificial-intelligence ecosystem: how many public repositories expose a license, a description, or a dependency specification, and how they are maintained. This gap matters because a model published without these artifacts cannot be legally reused, installed, or rebuilt, whatever its reported accuracy, and because clinicians who increasingly encounter open-source dental tools in their daily practice need to know whether those tools can be trusted and audited before they are relied upon. The objective of this study was to characterize the reproducibility of open-source dental artificial-intelligence software hosted on GitHub by quantifying the proportion of repositories that provide a license, a description, a README, and a dependency specification, and by describing their maintenance through archiving status and star counts.

METHODS

We conducted a cross-sectional observational study of public GitHub repositories that implement artificial-intelligence methods for dental tasks. The data were collected on 11 April 2026 from the GitHub REST search application programming interface, version 3, using the repositories search endpoint described in the GitHub documentation, which returns public metadata for repositories whose name, description, or readme match the query terms.

Five English queries were submitted, namely “dental artificial intelligence”, “dental deep learning”, “dental machine learning”, “oral radiography segmentation”, and “dental caries detection model”, and the first 30 results of each query were retained in the descending-relevance order that the interface returns by default.

These queries were chosen to span the most common formulations used to describe dental artificial-intelligence software, from the general to the task-specific, so that the sample would reflect how a researcher or clinician would actually search for such tools, and each query was executed once at the same point in time so that the sample represents a snapshot of the ecosystem rather than a cumulative census.

GitHub was selected as the sampling frame because it is the dominant public host for scientific and machine-learning software, because its search interface exposes the metadata fields that the reproducibility framework of Heil et al.⁽¹³⁾ treats as essential, and because its contents endpoint allows the root files of a repository to be inspected without downloading the code.

Repositories were included when they were public and pertained to the development or evaluation of artificial-intelligence software for a dental or oral task, and duplicate entries retrieved by more than one query were collapsed to a single record by matching the full repository name. No repository was excluded on the grounds of programming language, maturity, license, or star count, because the aim was to describe the ecosystem as it exists rather than a curated subset of exemplary projects. The final analytical sample comprised 78 repositories.

For every included repository we extracted the following metadata from the search response: the presence and identifier of a license, the presence of a text description, the primary programming language declared by the repository owner, whether the repository was archived, and its number of stars as a continuous measure of community usage.

The license, description, language, and archived fields are provided directly by the interface, and the star count was recorded as reported by the same endpoint at the time of retrieval. To assess the reproducibility of the software itself rather than only its discoverability, we then selected the 30 repositories with the highest star counts and queried the contents endpoint for each one, which lists the files and directories at the repository root.

From these listings we recorded whether a README file was present under any filename containing the token “readme”, and whether a dependency specification was present, defined as any root-level file named requirements.txt, environment.yml, pyproject.toml, or setup.py. The distinction between whole-sample metadata and file-level inspection follows the reporting dimensions advanced by Heil et al.,⁽¹³⁾ who proposed that machine-learning studies be reproducible by providing code, data, and environment descriptions, and by Wilkinson et al.,⁽¹⁴⁾ whose FAIR principles treat a license as a condition for reuse and a documented environment as a condition for reproducibility.

The operational definitions were fixed before extraction to avoid decisions that could be shaped by the data. A repository was considered to have a license when the license field returned a key rather than a null value, and license identifiers were grouped into standard open-source licenses,

such as the MIT license, the Apache license, the GNU General Public License, and the Creative Commons Zero dedication, versus non-standard or unrecognized identifiers.

A repository was considered to have a description when the description field contained non-whitespace text, and a README was considered present when a file whose name contained the token “readme” appeared at the repository root, irrespective of its extension or capitalization. A dependency specification was considered present only for the four named files, because those are the conventional mechanisms through which Python-based projects, the dominant language in this domain, declare the packages needed to run the code. Python was recorded as the primary language when the interface returned that value, and Jupyter Notebook was recorded separately because it signals a collection of notebooks rather than a package. These definitions mirror the checklist items that Schwendicke et al.⁽¹⁵⁾ recommended for reporting dental artificial-intelligence studies, adapted here from the level of a single study to the level of a repository.

The analysis was descriptive. We computed the absolute number and percentage of repositories carrying each metadata or file item, and we summarized the star counts with the median, the minimum, the maximum, and the mean because the distribution of stars is skewed and a median is more informative than an average. Percentages for the metadata items were computed over the full sample of 78 repositories, and percentages for the file items were computed over the 30 most popular repositories, because only those repositories were inspected at file level. The descriptive tables were generated directly from the computed metrics, and the percentage values were rounded to one decimal place for reporting, so that the figures in the text and tables agree to the precision shown.

All extraction and computation steps were performed by a single Python 3.11 script that read the saved repository records and the contents responses, so that the results are themselves reproducible from the raw data, and no manual transcription of numbers was required; this mirrors the very reproducibility principle that the study sets out to measure in the repositories themselves. No statistical hypothesis test was required, since the study aimed to estimate prevalence rather than to compare groups or test associations. Because the data are public, anonymized repository metadata and no human participants or patient records were involved, no ethical approval was sought; this treatment is consistent with the use of public software repositories as non-human administrative data, as is conventional in software-engineering studies. The raw repository records and the analysis script were retained in the working directory so that the reported figures can be recomputed.

RESULTS

Across the 78 repositories retrieved, the metadata that would permit legal reuse was the exception rather than the rule. A license was present in 23 of 78 repositories (0.3%), and the licensed minority was itself dominated by permissive and copyleft licenses: the MIT license was the most common (12 repositories), followed by the Apache license (4 repositories), while the GNU General Public License, the Affero General Public License, the Creative Commons Zero dedication, and the BSD three-clause license appeared in one repository each; 21 repositories carried a standard open-source license and 2 carried a non-standard or unrecognized identifier (Table 1).

A text description was present in 74 repositories (0.9%), so that most projects could be located and understood even when they could not be legally reused. The primary language was Python for 17 repositories (0.2%) and Jupyter Notebook for 30 (30 of 78), while 14 repositories (14 of 78) declared no language at all and the remaining 17 used other languages such as HTML, TypeScript, or JavaScript. None of the repositories (0) had been archived by its owner.

Table 1. Reproducibility metadata across 78 open-source dental AI repositories.

Reproducibility item	n	percent
License present	23	0.3%
Description present	74	0.9%
Python language	17	0.2%
Jupyter Notebook language	30	30.0%
Archived repository	0	0.0%

When the 30 most popular repositories were inspected at file level, documentation was near-universal but the environment needed to reproduce the software was not. All 30 of the 30 repositories (1.0%) contained a README file at the repository root, whereas only 8 (0.3%) provided a dependency specification in the form of requirements.txt, environment.yml, pyproject.toml, or setup.py (Table 2).

Table 2. File-level checks across the 30 most popular repositories.

Item (top 30 repos)	n	percent
README present	30	1.0%
Dependency file present	8	0.3%

This means that a reader who located one of these popular repositories could almost always learn what the project did and how to start it, but in only one quarter of cases could learn which exact packages and versions were required to run it without guessing.

Maintenance, as reflected in community usage, was modest and skewed. The median repository attracted 6 stars, and the mean was 16.40, with a range from 0 to 169; 12 repositories had received no stars at all, 22 had received ten or more, and only 3 had reached one hundred or more. The mean exceeded the median, a sign of right skew in which a few highly starred repositories pull the average upward while the majority cluster near zero.

The absence of archived repositories therefore coexisted with a long tail of low-star projects, which indicates that most open-source dental artificial-intelligence software is small in reach and has not been formally discontinued, but also that the community signal of endorsement is concentrated in a handful of projects while the majority remain unused or unvetted.

DISCUSSION

The central finding of this study is that open-source dental artificial-intelligence software is, for the most part, not legally reusable and not readily reproducible: fewer than one third of the 78 repositories carried a license, and only one quarter of the most popular repositories specified their dependencies.

This pattern is consistent with the reproducibility crisis that Hutson(7)described for artificial intelligence generally, and it extends the observation of Beam, Manrai and Ghassemi(11)that health-care machine-learning models rarely report what is needed for verification. Unlike the models reviewed by McDermott et al.,(12)whose failures were dominated by incomplete reporting of data and code, the repositories examined here were, by definition, open in their code; what they lacked was the legal and technical scaffolding that turns available code into a reproducible artifact. The finding therefore reframes the problem for dentistry: openness of source does not by itself deliver reproducibility, because a repository can be public and yet unlicensed and uninstallable.

The near-universal presence of a README among the most popular repositories is the most encouraging result, and it is worth setting against the scarcity of licenses. A README tells a reader what a project does, but it does not grant permission to use it; as Wilkinson et al.(14)pointed out, reuse depends on a clear license, and without one a repository defaults to restrictive copyright.

The contrast we observed, 1.0% with a README against 0.3% with a license, suggests that dental artificial-intelligence developers have internalized documentation but not the legal dimension of open science. Along the same lines, Heil et al.(13)treated the sharing of a runnable environment as a reproducibility standard, and the fact that only 0.3% of the most popular repositories declared their dependencies means that even a motivated reader must reconstruct the software stack by trial and error before the code can run.

The low proportion of repositories that declare a dependency specification has a concrete consequence for the clinical field that Schwendicke, Samek and Krois(1)envisioned. A dental clinician or researcher who downloads a caries-detection model but cannot install its dependencies cannot validate the model on local radiographs, and therefore cannot move from reported accuracy to trustworthy deployment. Schwendicke et al.(15)made the reporting of such implementation details a checklist item precisely because reviewers cannot otherwise judge a model, and Brinz et al.(17) observed that responsible data sharing in dentistry depends on the same practical infrastructure. Jansen et al.(19)and Jebari et al.(18)showed, in the life sciences and in open-source machine-learning

projects generally, that containerization and dependency declaration are feasible and increasingly common, which suggests that the dental community could adopt these practices without technical obstacles and that the gap is one of habit rather than capability.

The finding that a description was present in 0.9% of repositories while a license was present in only 0.3% points to a mismatch between discoverability and reusability that the FAIR principles of Wilkinson et al.(14) anticipate. A description and a README make a repository findable and interpretable, but reusability requires a license and a reproducible environment; the ecosystem we measured is therefore findable without being reusable.

This pattern is not unique to dentistry. Attafi et al.(20) reported that even in biology, where reporting standards are more mature, supervised machine-learning studies often omit the information that would make them reusable, and Shao et al.(21) identified the same gap in image classification. The consistency of these observations across domains indicates that the shortfall is not a failure of dental developers alone but a general weakness of applied machine-learning communities, a pattern that may reflect how rarely journals or hiring committees reward reproducibility work compared with novel accuracy results.

The language distribution offers a further clue to the reproducibility shortfall. Jupyter Notebook was the most common declared language, appearing in 30 repositories, more than the 17 repositories declared as Python packages.

Notebooks are convenient for demonstration but are difficult to reproduce, because they interleave code, narrative, and output in a way that hides the execution environment and encourages out-of-order cell execution; Kapoor and Narayanan(10) observed that even apparently clean notebooks can conceal non-reproducible steps, and Hutson(8) noted that model results presented without the means to rerun them are closer to demonstration than to science.

The preponderance of notebooks over packages therefore suggests that much dental artificial-intelligence software is written for display rather than for reuse, which compounds the effect of the missing dependency specifications and helps explain why documentation is high while reproducibility is low.

The results also clarify a terminological point that matters for how the field describes its own tools. A repository that is publicly visible on GitHub is often described as open source even when it carries no license, but in legal terms a repository without a license remains under default copyright and cannot be freely used, modified, or redistributed.

The finding that only 0.3% of repositories carried a license therefore means that a majority of what the community calls open-source dental artificial intelligence is not, in a strict sense, open source at all. This distinction was central to the FAIR principles of Wilkinson et al.,(14) and it has practical consequences for citation and collaboration, because researchers who build on unlicensed code may later discover that their use was not permitted, and clinicians who deploy such code may find that they lack the right to audit or maintain it.

The modest median of 6 stars and the complete absence of archived repositories describe an ecosystem that is small, young, and not yet formally abandoned. Just as Abd Rahim et al.(6) found that dental artificial intelligence remains in an early translational phase, the star distribution we observed suggests that most projects have not attracted broad adoption and that maintenance rests on a few active maintainers whose priorities may change.

Rokhshad et al.(16) warned that reliance on such software in clinical decisions carries ethical weight, because a model that is abandoned or unverifiable cannot be audited. Unlike the federated approaches that Rischke et al.(22) proposed as a way to train models without pooling patient data, open-source distribution leaves the burden of verification on the end user, and that burden is heavier when the repository lacks a license and a dependency specification, because the user cannot lawfully adopt the model and cannot easily run it.

Several specific applications illustrate what is at stake. Bittar, Dascanio and Özcan (23) and Lai et al.(24) reviewed artificial intelligence for caries detection and found large numbers of models with strong reported performance, and Horvath, Gjerdet and Shi (25) reached a similar conclusion for bitewing radiography.

Cui et al.(26) published a panoramic-radiography dataset, and Aydin and Can (27) described the advances and remaining challenges of deep learning in dental imaging. What none of these contributions can ensure is that the underlying repositories will remain installable and reusable over time, and our data suggest that a substantial share would not be. Suganya et al.(28) added that fairness and equity in dental artificial-intelligence systems cannot be assessed when the model cannot be inspected, which reinforces the reproducibility requirement rather than competing with it: a model that cannot be audited is also a model whose bias cannot be known.

The growth documented by these authors therefore overstates the practical availability of

dental artificial intelligence, because the literature reports results while the software that produced them remains, in many cases, legally and technically closed despite being nominally open source.

On the basis of these results, we suggest three low-cost practices for the dental artificial-intelligence community. First, every public repository should carry a permissive license, because this is the single change that most improves legal reusability and it requires only one metadata field. Second, every repository should declare its dependencies through a requirements file or an equivalent, because this is what separates code that can be run from code that can only be read. Third, journals and reviewers that evaluate dental artificial-intelligence manuscripts should ask authors to provide both, following the checklist of Schwendicke et al.,⁽¹⁵⁾ so that publication and reproducibility advance together rather than in opposition.

The practical implication for the dental community is that discoverability is not a sufficient condition for adoption. A clinician who locates a repository through a search engine has, in most cases, no legal permission to use it and no specification of the environment needed to run it, so the current open-source ecosystem delivers the appearance of availability without its substance. This gap also has consequences for journals and reviewers: as Schwendicke et al.⁽¹⁵⁾ proposed, manuscripts reporting artificial-intelligence models should be evaluated on the reproducibility of their software, and our results suggest that a large share of the software that dental researchers would cite or reuse would fail such an evaluation.

Attafi et al.⁽²⁰⁾ made the same argument for biology through the DOME registry, and the dental community could adopt an analogous expectation that authors deposit a license and a dependency specification alongside their code. Because the infrastructure for doing so already exists and is widely used in other domains, as Jebari et al.⁽¹⁸⁾ demonstrated, the barrier is procedural rather than technical, and a small change in developer habit would measurably improve the reusability of the ecosystem.

The study has several strengths that support the interpretation of these results. It is, to our knowledge, the first systematic description of the open-source dental artificial-intelligence ecosystem measured directly from the hosting platform rather than inferred from published papers, and it therefore avoids the publication bias that affects literature-based reviews such as those of Singh and Raza⁽³⁾ and Sivari et al.,⁽⁴⁾ which describe only the models that authors chose to report.

The operational definitions were fixed before extraction and applied by a single script, which removes coder variability, and the use of the GitHub interface with explicit, reproducible queries makes the study transparent and repeatable by any reader. The combination of whole-sample metadata with file-level inspection of the most popular repositories provides both breadth across the ecosystem and depth on the projects that are most likely to be adopted, which is the appropriate compromise between completeness and feasibility.

The study has limitations that follow from its design, and each suggests a specific remedy. The file-level inspection was limited to the 30 most popular repositories, so the prevalence of README and dependency files among the long tail of low-star projects remains unknown; extending the file inspection to the full sample would test whether documentation and license coverage are poorer still outside the most visible projects, which would likely lower the estimates we report. The absence of an archived repository does not confirm active maintenance, and a star count is only a proxy for use rather than a measure of correctness; longitudinal re-examination would distinguish maintained from silently dormant projects. Finally, the study measured the presence of reproducibility artifacts but not whether a model actually reproduces when rebuilt; as Dash and Patil⁽²⁹⁾ and Straiton⁽³⁰⁾ argued, execution-based verification is the stronger test and would complement the static assessment reported here by attempting to install and run a sample of the repositories.

CONCLUSIONS

Open-source dental artificial-intelligence software was characterized by weak legal reusability, because fewer than one third of repositories carried a license, and by weak technical reproducibility, because only one quarter of the most popular repositories specified their dependencies, despite near-universal README documentation and a prevalence of notebook-based over package-based projects.

The ecosystem was small and unarchived, with a median of six stars and a long tail of unused projects, which indicated early, concentrated, and as-yet unformalized maintenance and a field still clearly in the process of consolidation rather than maturity.

These results imply that developers should attach a permissive license and a dependency specification as a default, that journals and reviewers should require those artifacts before endorsing

a model, and that clinicians should treat their absence as a warning when adopting open-source software for dental use, because a repository that cannot be lawfully reused or reliably run offers little more than an illustration.

REFERENCES

- Schwendicke F, Samek W, Krois J. Artificial Intelligence in Dentistry: Chances and Challenges. *J Dent Res*. 2020;99(7):769-774. doi: 10.1177/0022034520915714
- Katsumata A. Deep learning and artificial intelligence in dental diagnostic imaging. *Japanese Dental Science Review*. 2023;59:329-333. doi: 10.1016/j.jdsr.2023.09.004
- Singh NK, Raza K. Progress in deep learning-based dental and maxillofacial image analysis: A systematic review. *Expert Systems with Applications*. 2022;199:116968. doi: 10.1016/j.eswa.2022.116968
- Sivari E, Senirkentli GB, Bostanci E, Guzel MS, Acici K, Asuroglu T. Deep Learning in Diagnosis of Dental Anomalies and Diseases: A Systematic Review. *Diagnostics*. 2023;13(15):2512. doi: 10.3390/diagnostics13152512
- AbuSalim S, Zakaria N, Islam MR, Kumar G, Mokhtar N, Abdulkadir SJ. Analysis of Deep Learning Techniques for Dental Informatics: A Systematic Literature Review. *Healthcare*. 2022;10(10):1892. doi: 10.3390/healthcare10101892
- Abd Rahim IN, Omar N, Zakaria N, Zulkapli R, Ahmad R, Kamar Affendi NH, et al. Current status, evidence maturity, and translational readiness of artificial intelligence in dentistry: a bibliometric and thematic analysis. *BDJ Open*. 2026;12(1):89. doi: 10.1038/s41405-026-00478-1
- Hutson M. Artificial intelligence faces reproducibility crisis. *Science*. 2018;359(6377):725-726. doi: 10.1126/science.359.6377.725
- Hutson M. Has artificial intelligence become alchemy?. *Science*. 2018;360(6388):478-478. doi: 10.1126/science.360.6388.478
- Gibney E. Could machine learning fuel a reproducibility crisis in science?. *Nature*. 2022;608(7922):250-251. doi: 10.1038/d41586-022-02035-w
- Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*. 2023;4(9):100804. doi: 10.1016/j.patter.2023.100804
- Beam AL, Manrai AK, Ghassemi M. Challenges to the Reproducibility of Machine Learning Models in Health Care. *JAMA*. 2020;323(4):305. doi: 10.1001/jama.2019.20866
- McDermott MBA, Wang S, Marinsek N, Ranganath R, Foschini L, Ghassemi M. Reproducibility in machine learning for health research: Still a ways to go. *Sci Transl Med*. 2021;13(586):eabb1655. doi: 10.1126/scitranslmed.abb1655
- Heil BJ, Hoffman MM, Markowetz F, Lee SI, Greene CS, Hicks SC. Reproducibility standards for machine learning in the life sciences. *Nat Methods*. 2021;18(10):1132-1135. doi: 10.1038/s41592-021-01256-7
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3(1):160018. doi: 10.1038/sdata.2016.18
- Schwendicke F, Singh T, Lee JH, Gaudin R, Chaurasia A, Wiegand T, et al. Artificial intelligence in dental research: Checklist for authors, reviewers, readers. *Journal of Dentistry*. 2021;107:103610. doi: 10.1016/j.jdent.2021.103610
- Rokhshad R, Ducret M, Chaurasia A, Karteva T, Radenkovic M, Roganovic J, et al. Ethical considerations on artificial intelligence in dentistry: A framework and checklist. *Journal of Dentistry*. 2023;135:104593. doi: 10.1016/j.jdent.2023.104593
- Brinz J, Eslamiamirabadi N, Salamati A, Tresp V, Schwendicke F, Tichy A. Data sharing for responsible artificial intelligence in dentistry: a narrative review of legal frameworks and privacy-preserving techniques. *Journal of Dentistry*. 2025;163:106130. doi: 10.1016/j.jdent.2025.106130
- Jebari F, Ksontini E, Barrak A, Kessentini W. ML in a Box: Analyzing Containerization Practices in Open Source ML Projects. *Proceedings of the 23rd International Conference on Mining Software Repositories*. 2026:122-132. doi: 10.1145/3793302.3793344
- Jansen C, Annuscheit J, Schilling B, Strohmenger K, Witt M, Bartusch F, et al. Curious Containers: A framework for computational reproducibility in life sciences with support for Deep Learning applications. *Future Generation Computer Systems*. 2020;112:209-227. doi: 10.1016/j.future.2020.05.007
- Attafi OA, Clementel D, Kyritsis K, Capriotti E, Farrell G, Fragkouli SC, et al. DOME Registry: implementing community-wide recommendations for reporting supervised machine learning in biology. *GigaScience*. 2024;13:giae094. doi: 10.1093/gigascience/giae094
- Shao G, Zhang H, Shao J, Woeste K, Tang L. Strengthening Machine Learning Reproducibility for Image Classification. *AAIML*. 2022;02(04):471-476. doi: 10.54364/aaaiml.2022.1132
- Rischke R, Schneider L, Müller K, Samek W, Schwendicke F, Krois J. Federated Learning in Dentistry: Chances and Challenges. *J Dent Res*. 2022;101(11):1269-1273. doi: 10.1177/00220345221108953
- Bittar A, Dascanio R, Özcan M. Artificial intelligence for dental caries detection: An umbrella review. *Journal of Dentistry*. 2026;175:106944. doi: 10.1016/j.jdent.2026.106944
- Lai J, Guo S, Wang K, Li X, Yu X, Wang X. Artificial intelligence for binary dental caries diagnosis using intraoral images and dental radiographs: a systematic review and meta-analysis. *BMC Oral Health*. 2026;26(1):437. doi: 10.1186/s12903-026-07770-4
- Horvath KSH, Gjerdet NR, Shi XQ. Advancements in Caries Diagnostics Using Bitewing Radiography: A Systematic Review of Deep Learning Approaches. *Caries Res*. 2025;60(2):127-151. doi: 10.1159/000546448
- Cui J, Gu J, Guan Y, Xiao H, Liu K, Zhang W, et al. A multi-focus oral panoramic x-ray image dataset based on pixel-level annotations. *Sci Data*. 2026;13(1):693. doi: 10.1038/s41597-026-07021-9
- Aydin E, Can Z. Deep Learning in Dental Imaging: Advances, Challenges, and Future. *Oral Radiol*. 2026;42(4):853-879. doi: 10.1007/s11282-026-00939-1

28. Suganya P, Dupada P, Sruthi KG, Panda P, Mohanty JR. Bias, fairness, and equity in artificial intelligence systems used in dental imaging: A systematic review. *International Journal of Medical Informatics*. 2026;214:106433. doi: 10.1016/j.ijmedinf.2026.106433
29. Dash S, Patil K. A Reproducible Experimental Protocol for Developing and Evaluating Explainable Artificial Intelligence Models in Healthcare Systems. *JoVE*. 2026;(235). doi: 10.3791/73848
30. Straiton J. Artificial intelligence: help or hindrance in solving the reproducibility crisis?. *BioTechniques*. 2024;76(7):291-294. doi: 10.1080/07366205.2024.2355776

FUNDING

No financing

CONFLICT OF INTEREST

None

AUTHORSHIP CONTRIBUTIONS

Drafting – original draft: Flor M Escalante Freitez, Roberto Antonio Vega Bonilla, Diana Belkis Moreno Benítez.

Writing–review and editing: Flor M Escalante Freitez, Roberto Antonio Vega Bonilla, Diana Belkis Moreno Benítez.