

Article

Not peer-reviewed version

Integrating Machine Learning and Deep Learning for Predicting Non-Surgical Root Canal Treatment Outcomes Using Two-Dimensional Periapical Radiographs

[Catalina Bennasar](#)*, [Antonio Nadal-Martínez](#), [Sebastiana Arroyo](#), [Yolanda Gonzalez](#), [Ángel Arturo López-González](#), [Francesc Pérez](#), [Pedro Juan Tárraga](#)

Posted Date: 4 March 2025

doi: 10.20944/preprints202503.0275.v1

Keywords: machine learning; deep learning; outcome prediction; non-surgical root canal treatment; apical periodontitis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Integrating Machine Learning and Deep Learning for Predicting Non-Surgical Root Canal Treatment Outcomes Using Two-Dimensional Periapical Radiographs

Catalina Bennasar ^{1,*}, Antonio Nadal ², Sebastiana Arroyo ³, Yolanda Gonzalez-Cid ⁴, Ángel Arturo López-González ⁵, Francesc Pérez ⁶ and Pedro Juan Tárraga ⁷

¹ ADEMA, School of Dentistry, University of Balearic Islands, 07122 Palma de Mallorca, Spain

² Soft Computing, Image Processing and Aggregation (SCOPIA) Research Group, University of the Balearic Islands (UIB), 07122 Palma, Spain

³ Coordinating teacher of Conservative Dentistry, ADEMA School of Dentistry, Re-searcher of ADEMA Health IUNICS group, University of the Balearic Islands. Palma, Balearic Islands, Spain, Aggregate professor of Barcelona University, IDIBELL Researcher

⁴ Department of Mathematical Sciences and Informatics, University of the Balearic Islands, 07120 Palma de Mallorca, Spain

⁵ ADEMA-Health group, University Institute of Health Sciences of Balearic Islands (IUNICS), Palma de Mallorca, Spain

⁶ Dental Public Health Service, IB-Salut, Balearic Islands, 07003 Palma de Mallorca, Spain

⁷ Faculty of Medicine. University of Castilla-La Mancha. Albacete. Spain

* Correspondence: c.bennasar@eua.edu.es; Tel.: +34-626-388-003

Abstract: Background/Objectives: In a previous study, we utilized categorical variables and Machine Learning (ML) algorithms to predict the success of non-surgical root canal treatments (NSRCTs) in apical periodontitis (AP), classifying the outcome as either success (healed) or failure (not healed). Given the importance of radiographic imaging in diagnosis, the present study evaluates the efficacy of Deep Learning (DL) in predicting NSRCT outcomes using two-dimensional (2D) periapical radiographs, comparing its performance with ML models. **Methods:** The DL model was trained and validated using Leave-One-Out Cross-Validation (LOOCV). Its output was incorporated into the set of categorical variables, and the ML study was reproduced using Backward Stepwise Selection (BSS). The chi-square test was applied to assess the association between this new variable and NSRCT outcomes. Finally, after identifying the best-performing method from the ML study reproduction, statistical comparisons were conducted between this method, clinical professionals, and the image-based model using Fisher's Exact Test. **Results:** The association study yielded a p-value of 0.000000127, highlighting the predictive capability of 2D radiographs. After incorporating the DL-based predictive variable, the ML algorithm that demonstrated the best performance was Logistic Regression (LR), differing from the previous study, where Random Forest (RF) was the top performer. When comparing the Deep Learning – Logistic Regression (DL-LR) model with the clinician's prognosis (DP), DL-LR showed superior performance with a statistically significant difference (p-value < 0.05) in sensitivity, NPV, and accuracy. The same trend was observed in the DL vs. DP comparison. However, no statistically significant differences were found in the comparisons of RF vs. DL-LR, RF vs. DL, and DL vs. DL-LR. **Conclusions:** The findings of this study suggest that image-based artificial intelligence models exhibit superior predictive capability compared to those relying exclusively on categorical data. Moreover, they outperform clinician prognosis.

Keywords: machine learning; deep learning; outcome prediction; non-surgical root canal treatment; apical periodontitis

1. Introduction

Over the past decade, ML and DL have emerged as transformative technologies with significant impacts across multiple scientific disciplines. These techniques, as branches of artificial intelligence (AI), have become essential tools for analyzing large volumes of data, identifying complex patterns, and providing innovative solutions to previously intractable problems. Their relevance has been particularly emphasized in health sciences, including medicine, biology, and more recently, dentistry, where their application holds the potential to optimize diagnostics, personalize treatments, and improve clinical outcomes.

ML can be defined as a subfield of AI that employs algorithms to enable machines to learn patterns and behaviors from data without being explicitly programmed for each specific task (Samuel, 1959). Within this field, DL represents a significant evolution, utilizing artificial neural networks with multiple layers capable of learning hierarchical data representations. These deep architectures have proven especially effective in tasks involving large datasets and complex features, such as medical image interpretation, disease prediction, and biological signal analysis.

The current utility of ML and DL in health sciences is demonstrated in a wide range of applications, including but not limited to medical image analysis, genomic data mining, drug modeling, and clinical outcome prediction. For instance, in radiology, convolutional neural networks (CNNs) have been successfully employed to detect tumor lesions in computed tomography (CT) and magnetic resonance imaging (MRI) scans, achieving accuracy levels comparable to those of human experts [1]. In genomics, deep learning algorithms have facilitated the decoding of complex gene interactions, accelerating the development of personalized therapies [2].

In the field of dentistry, the impact of ML and DL is beginning to solidify with promising applications. Dentistry, as a health science, has undergone significant digital transformation in recent years, driven by technologies such as cone-beam computed tomography (CBCT), three-dimensional (3D) printing, and CAD/CAM systems. However, the integration of ML and DL in this domain has opened new avenues for diagnosis, treatment design, and disease monitoring. For example, deep learning algorithms have been effective in identifying dental caries [3], fractures, periodontal diseases [4], and periapical conditions from digital radiographs, enhancing diagnostic accuracy and reducing variability among professionals [5].

Beyond diagnostics, these technologies are beginning to influence treatment planning and execution. In orthodontics, for example, ML is used to predict tooth movement and optimize the placement of brackets or aligners, leading to more effective and personalized treatments [6]. In implantology, ML models assist in predicting dental implant stability over time by considering factors such as bone density, implant location, and patient characteristics [7]. In endodontics, AI supports professionals by detecting periapical lesions, identifying root fractures, analyzing root canal morphology, predicting retreatment needs, and aiding in regenerative pulpal therapy, all of which contribute to improved diagnostics, treatment planning, and patient care [8–13].

The future of ML and DL in health sciences, particularly in dentistry, promises to be even more revolutionary. It is anticipated that the combination of these technologies with advanced sensor systems and data from wearable devices will enable continuous, real-time monitoring of oral health. For example, the integration of DL with intraoral connected devices could facilitate the early detection of diseases such as oral cancer through the analysis of salivary biomarkers or intraoral images [14]. Additionally, the development of explainable AI (XAI) systems could address one of the most pressing current challenges: the need to provide clear and interpretable explanations of algorithm-generated predictions, fostering clinical acceptance and ethical use [15].

Despite these promises, the implementation of ML and DL in clinical practice faces significant challenges that must be addressed to ensure success. These include the need for large volumes of high-quality data for model training, ethical and legal concerns related to data privacy, and the necessity of educating healthcare professionals in the use of these technologies. These challenges highlight the importance of interdisciplinary collaboration involving researchers, technology developers, clinicians, and policymakers [16,17].

2. Study Objectives

- To review the current state of ML and DL in health sciences, with a special focus on dentistry, highlighting their current applications, future opportunities, and the barriers that must be overcome to maximize their impact. By exploring these aspects, we hope to provide a comprehensive perspective that contributes to advancing this fascinating intersection of technology and clinical innovation.
- To evaluate DL as an additional variable in a ML study for predicting NSRCTs in cases of AP. The study aims to determine the extent to which deep neural networks can predict the outcome of NSRCTs in teeth with apical periodontitis using digital periapical radiographs of confirmed AP diagnoses.

3. Materials and Methods

3.1. Sample Selection

A retrospective study was conducted based on the analysis of clinical records of patients with AP who underwent NSRCTs for the first time (not re-treatments). Cases were randomly selected from the database of a private clinic in Mallorca, Spain. Only patients without reported systemic diseases who received treatment for the first time and whose records included the following were included:

A comprehensive medical and dental history with general, facial, and oral inspection reports, as well as dental percussion and palpation examinations;

Results of complementary tests, such as thermal sensitivity testing using an ice pencil and periapical radiography;

A follow-up period of at least nine years, starting six months after treatment, with documented evaluations of lesion recovery, categorizing cases as successful (0: there are no symptoms or indications for further treatments, and the lesion disappears after NSRCT) or failed (1: the failure occurs when either the clinical or radiographic outcome fails).

Radiographs were acquired using an X Mind Unity Acteon Satelec system with a focal point of 0.4 mm, at 70 kV and 7 mA, employing a Carestream 6100 digital system with a resolution of 15 LP/mm. The bisecting angle technique was used with a Rinn XCD (Dentsply) positioner. Patients with vertical root fractures or teeth without sufficient ferrule structure for subsequent restoration were excluded.

Due to this filtering process, the final number of patients included in the study was reduced to 119. Patient consent was waived due to the inability to identify participants in the database. The Research Ethics Committee of the Balearic Islands (IB4015/19IP) approved the study.

3.2. Intervention Procedure

The 119 patients with confirmed AP, for whom eight preoperative domain variables were observed as per a recommended data collection template (DCT) for endodontic treatment evaluation studies [18–20], underwent standardized endodontic treatment performed by the same endodontist using identical materials and procedures. The following phases were followed:

Local anesthesia administration and rubber dam placement.

Chamber access and pre-enlargement of the coronal third, followed by apical third negotiation.

Working length determination using a Morita apex locator and radiographic confirmation. The working length was always set at the radiographic apex level.

Instrumentation with K3 (SybronEndo) and Protaper Gold (Dentsply Maillefer) rotary systems, complemented with manual instruments.

Irrigation with EDTA and 5.25% sodium hypochlorite.

Obturation using the warm vertical condensation technique with AH Plus sealer.

Following treatment completion, cases were radiographically evaluated to rule out overfills or obturation defects.

3.3. Machine Learning and Deep Learning Analysis

To compare DL with ML models, a previous study titled "Second Opinion for NSRCT Prognosis Using Machine Learning Models" [21] was utilized, where Logistic Regression (LR), Random Forest (RF), Naive Bayes (NB), and K-Nearest Neighbors (KNN) algorithms were applied. The RF algorithm demonstrated the best performance.

For DL-based analysis, diagnostic radiographs were exported to a database with identifiers and labels (0: healed, 1: not healed) (Figures 1 and 2).

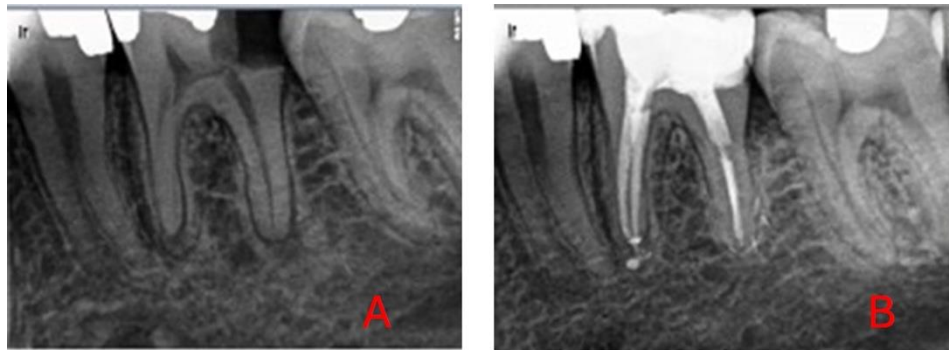


Figure 1. A: AP in tooth 36. B: Three-year follow-up: healed.

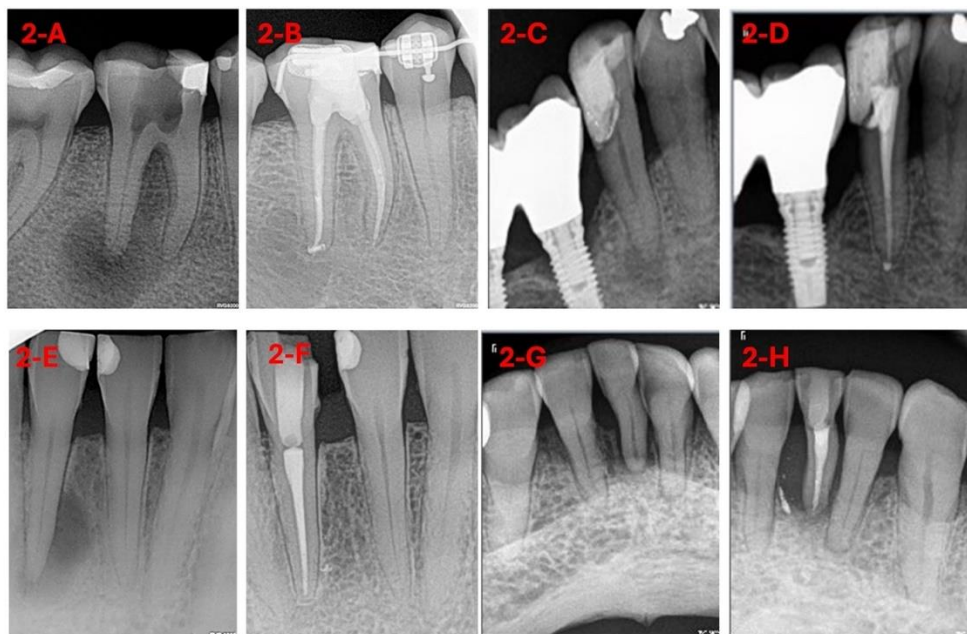


Figure 2. 2-A: AP in tooth 46. 2-B: One-year follow-up: healed. 2-C: AP in tooth 45. 2-D: Four-year follow-up: healed. 2-E: AP in tooth 41. 2-F: Three-year follow-up: healed. 2-G: AP in tooth 31. 2-H: Two-year follow-up: not healed.

The AnotIA software was utilized for the precise segmentation of diagnostic 2D periapical radiographic images of AP, assigning labels to facilitate subsequent analysis (Figure 3).

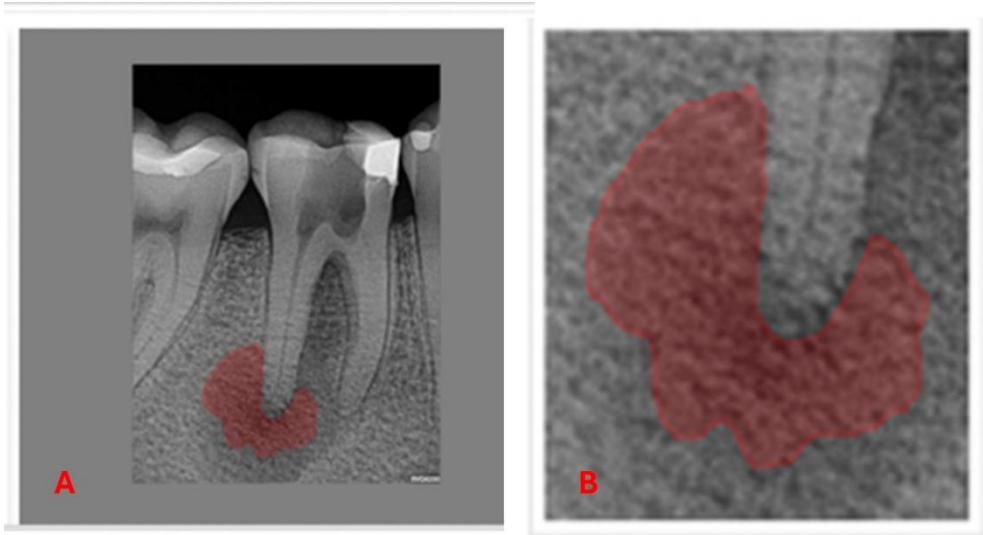


Figure 3. Demarcation of apical periodontitis.

Diagnostic two-dimensional AP images were employed to train a convolutional neural network based on the ResNet-18 architecture, a deep 18-layer network designed for recognizing complex patterns in medical images. This architecture has demonstrated efficacy in various AI applications in dentistry [22–27] due to its residual connections, which facilitate deep network training and mitigate accuracy degradation issues.

This study aims to provide evidence regarding the predictive capability of DL in NSRCT prognosis, comparing its performance with conventional ML models and validating its applicability in clinical settings.

4. Statistical Analysis and Results

For statistical analysis, we relied on the results obtained in our previous study, where a set of preoperative patient variables, both clinical and demographic, were used as explanatory covariates in various ML models to predict treatment outcomes [21]. In the present study, we included an additional explanatory covariate: the treatment outcome prediction obtained by applying convolutional networks to the diagnostic images of 108 patients, training the networks to forecast the prognosis.

Using the results obtained from the DL study, we applied the chi-square test to detect the association between DL results (DL prediction) and the dentist’s outcome, obtaining a p-value of 0.000000127 and an Effect Size of 0.53 (Table 1).

Table 1. Variables associated with the results of the previous ML study, incorporating the DL prediction.

Variable	Levels	p-Value	Effect Size
Age	15-24; 25-34; 35-44; 45-54; 55-64; ≥65	0.0056	0.372
Highest level of education	Primary; Secondary; Post secondary	0.0016	0.33
Arch	Mandible; Maxilla	0.02	0.21
Smoking	No; Everyday; Someday; Former	0.046	0.26
Patient co-operation	No; Yes	0.028	0.21
Pain relieved by	None; Cold; Medication	0.003	0.31

Time-lasting of the pain	Sec; Min; Continuous	0.027	0.245
Periapical	Asymptomatic AP; Symptomatic AP; Chronic Apical Abscess; Acute Apical	0.01	0.31
Estimated Prognosis by	Hopeless; Questionable; Fair; Good;	0.034	0.29
Prediction by DL	Success; Failure	0.000000127	0.53

The training and validation process of DL follows the same LOOCV scheme used in the evaluation of ML algorithms, including Logistic Regression (LR), Random Forest (RF), Naive Bayes (NB), and K Nearest Neighbors (KNN) [21,28]. In this approach, the DL treatment prognosis for each patient is determined by training the model with the images of the remaining patients.

The use of LOOCV is particularly valuable for assessing the performance of artificial intelligence models, as it systematically excludes one data point from the training set, using it as a validation or test instance. Subsequently, a predictive value is generated for the excluded data, and this process is repeated as many times as elements are in the training set. Finally, the predicted values for each excluded data point are compared with the observed values, allowing for a rigorous evaluation of model performance.

In the ML study, once the variable "Prediction by DL" was incorporated into each of the models, the LOOCV scheme was applied again. For variable selection, the Backward Stepwise Selection (BSS) technique was used [28], a commonly employed method for identifying the most relevant features in predictive models. The best performance was obtained with a LR model, in which the most influential variables were: "DL" (predictions generated by DL networks), "Age", "Smoking", "Level_Education", "Periapical" (periapical condition), and "Prognosis".

The performance of all methods used in this study is presented in Table 2.

Table 2. "Performance of AI Algorithms and the Dentist Prognosis (DP)".

Metric	DP	RF	Logistic Regression (DL- LR)	DL
TP	42	57	57	59
FN	27	12	8	6
FP	21	15	15	18
TN	29	35	28	25
Sensitivity	0.61 (0.48, 0.72)	0.83 (0.72, 0.91)	0.87 (0.77, 0.94)	0.90 (0.80, 0.90)
Specificity	0.58 (0.43, 0.72)	0.7 (0.55, 0.82)	0.65 (0.49, 0.78)	0.58 (0.42, 0.72)
PPV	0.67 (0.54, 0.78)	0.79 (0.68, 0.88)	0.79 (0.67, 0.87)	0.76 (0.65, 0.85)
NPV	0.52 (0.38, 0.65)	0.74 (0.6, 0.86)	0.77 (0.60, 0.89)	0.80 (0.62, 0.92)
Accuracy	0.6 (0.5, 0.69)	0.77 (0.69, 0.84)	0.78 (0.69, 0.86)	0.77 (0.68, 0.85)

5. Statistical Comparisons

Having established the performance of DL and LR, after incorporating DL prediction as an additional explanatory covariate alongside patients' preoperative clinical and demographic variables used in ML models to predict the outcome of NSRCTs for AP [21], a series of statistical comparisons will be conducted.

For this analysis, Fisher's exact test will be employed, setting a significance level of 0.05. Any result with a p-value below 0.05 will be considered statistically significant. The comparisons to be evaluated are as follows:

- Comparison between the best ML model from the previous study (21), Random Forest (RF), and the combined Deep Learning and Logistic Regression model (DL-LR): (RF vs. DL-LR).
- Comparison between Random Forest and Deep Learning in general: (RF vs. DL).
- Comparison between the clinical professional's prediction (DP) and the combined Deep Learning and Logistic Regression model (DL-LR): (DP vs. DL-LR).
- Comparison between the clinical professional's prediction (DP) and the Deep Learning model in general (DL): (DP vs. DL).
- Comparison between the combined Deep Learning and Logistic Regression model (DL-LR) and the general Deep Learning model (DL): (DL-LR vs. DL).

These comparisons will assess the relative efficacy of different predictive approaches, providing valuable insights into the applicability of AI models in predicting the success of NSRCT

5.1. Comparison Between Random Forest (RF) and the Deep Learning-Logistic Regression Model (DL-LR)

Overall, DL-LR outperformed the best-performing Machine Learning model from the previous study [21], Random Forest, achieving sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy of 0.87, 0.65, 0.79, 0.77, and 0.78, respectively. In comparison, Random Forest yielded values of 0.83, 0.70, 0.79, 0.74, and 0.77 for the same metrics. However, the differences were not statistically significant, suggesting similar performance between the two models.

5.2. Comparison Between Random Forest and Deep Learning in General (RF vs. DL)

The comparative analysis between the overall Deep Learning model and Random Forest showed no statistically significant differences in their performance.

5.3. Comparison Between the Clinical Professional's Prediction (DP) and Deep Learning-Logistic Regression (DP vs. DL-LR)

In the comparison between DL-LR and DP, DL-LR demonstrated better performance in sensitivity, NPV, and accuracy. Using the true positive (TP) and false negative (FN) values from Table 2, a statistically significant difference was observed in the sensitivity of the Logistic Regression model with Deep Learning compared to the professional's prediction (p-value = 0.00041). However, using the false positive (FP) and true negative (TN) values from the same table, no significant differences were found in specificity or PPV. Nevertheless, significant differences were identified in NPV (p-value = 0.01563) and accuracy (p-value = 0.00253).

5.4. Comparison Between the Clinical Professional's Prediction (DP) and Deep Learning (DL) (DP vs. DL)

Similarly, when comparing the standalone Deep Learning model with DP, statistically significant differences were found in sensitivity (p-value = 0.00005), NPV (p-value = 0.0108), and accuracy (p-value = 0.00421), indicating superior performance of DL in these key metrics.

5.5. Comparison Between Deep Learning and the Combined Logistic Regression-Deep Learning Model (DL vs. DL-LR)

Finally, when comparing the individual Deep Learning (DL) model with the Logistic Regression model supplemented with categorical variables and the output of the DL model (DL-LR), no statistically significant differences were found in any of the evaluated metrics, suggesting equivalent performance between both models.

5.6. Interpretation of Statistical Comparisons

Based on statistical comparisons: DP vs. DL, RF vs. DL, and DL-LR vs. DL, the following conclusions can be drawn:

Categorical variables have a lower predictive value compared to image-based models.

Sensitivity, NPV, and accuracy metrics show minimal or non-significant differences between RF vs. DL and DL-LR vs. DL models.

High p-values (>0.05) in model comparisons indicate no real difference between categorical data-based approaches.

No clear improvements were found in any metric for RF vs. DL and DL-LR vs. DL comparisons.

In contrast, the DP vs. DL comparison showed significant differences:

DL vs. DP exhibited very low p-values (0.00005 in sensitivity, 0.00421 in accuracy, 0.0108 in NPV), suggesting that image-based models (likely represented by DL) have superior predictive power compared to dental professionals (DP).

Similarly, DP vs. AI-based methods showed significant differences:

DL-LR vs. DP displayed very low p-values (0.00041 in sensitivity, 0.00253 in accuracy, 0.01563 in NPV), indicating that AI-based methods (DL-LR) have better predictive value than dental professionals (DP).

6. Discussion

The results obtained in this study, supported by the statistical data collected, highlight the need to compare our AI-based NSRCT prediction for AP with existing literature to validate our findings scientifically. However, this comparison is challenging due to the limited number of studies dedicated to predicting NSRCT outcomes for apical periodontitis using AI applied to 2D periapical radiographs in endodontics.

AI systems have demonstrated significant advancements in medical imaging, substantially contributing to diagnosis and treatment planning across various specialties. In medicine, convolutional neural networks (CNNs) have been employed for the automatic analysis of pathologies such as breast cancer [29], lung cancer [30,31], and Alzheimer's disease [32]. In dentistry, AI applications have included dental caries detection [33], implant classification [34], periodontal bone loss quantification [35], and cyst evaluation using various types of radiographs, including periapical, panoramic, cephalometric, and CBCT images [36]. In endodontics, AI has been applied to detect apical periodontitis [5] and C-shaped root canals [37].

Although DL applications in medicine are well established [38], studies on disease and treatment outcome prediction in endodontics remain considerably limited [22,29,30]. In this context, Lee et al. (2023) [22] conducted a study predicting endodontic treatment and retreatment outcomes over a three-year period using 598 preoperative periapical radiographs of single-rooted premolars. Utilizing a ResNet-18 CNN model, trained, validated, and tested, their study focused on two main objectives: detecting various clinical features and predicting treatment outcomes. Their findings confirmed the feasibility of DCNN algorithms for feature detection and endodontic prognosis prediction.

Our study shares the objective of evaluating the predictive capability of endodontic treatments using DL with a ResNet-18 architecture; however, our methodology considers all tooth types, not just single-rooted premolars. The selection of single-rooted premolars in Lee et al.'s study [22] was based on the lower anatomical variability of these teeth compared to incisors or molars, which can present heterogeneous periapical conditions [39,40]. Additionally, all cases analyzed in our study exhibited AP, reducing treatment outcome variability. Unlike Lee et al.'s study [22], our research did not include retreatments, which can influence treatment success rates. Furthermore, our study's evaluation period was extended to nine years, whereas Lee et al. [22] conducted a three-year follow-up. This distinction is relevant, as short-term evaluations may not fully capture the healing process [40].

A key methodological aspect in endodontic treatment evaluation is the use of the Periapical Index (PAI) score. In Lee et al.'s study [22], only PAI scores 1, 4, and 5 were considered, omitting stage PAI 3, which reflects bone structural changes with minimal demineralization characteristic of apical periodontitis [41]. In our study, we opted to dichotomize the PAI evaluation to avoid ambiguities.

Moreover, our study accounted for working length and obturation type, critical parameters influencing treatment success rates [42–44].

In a broader context, the literature has explored various AI applications in endodontics. A study employing the AGMB-Transformer model used a dataset of 245 radiographic images of root canal treatments to evaluate its performance in anatomical structure segmentation and outcome classification [45]. Although this study did not focus on treatment prediction, it demonstrated that combining segmentation and classification data significantly improves automated evaluations.

Systematic reviews by Aminoshariae et al. [9] and Khanagar et al. [11] have consolidated knowledge on AI in endodontics, addressing areas such as diagnosis, clinical decision-making, and therapeutic success prediction. However, predicting endodontic treatment outcomes remains an unexplored research gap. Parvathi et al. [8] analyzed AI applications in endodontics, including apical foramen localization, root fracture detection, and retreatment prediction. Campo et al. [47] introduced a Case-Based Reasoning (CBR) system to minimize failed retreatments; however, literature addressing NSRCT outcome prediction remains scarce [48].

The use of ResNet-18 architectures in dentistry has proven to be an effective methodology for various applications, including dental caries classification [24], apical periodontitis detection [23], and periodontal disease evaluation [27]. Other studies have employed deep learning for anatomical structure segmentation [51], predicting inferior alveolar nerve paresthesia after third molar extraction [26], and detecting external root resorptions [25].

Despite advancements in AI applications in endodontics, current literature presents a shortage of studies focusing on predicting the outcomes of primary endodontic treatments for apical periodontitis. As evidenced by Lee et al. [22] and Yunxiang Li et al. [45], additional studies are imperative. Compared to medicine, where AI has demonstrated significant advancements, efforts in endodontics remain focused on detecting periapical lesions [13,16,23,48,55,57,58], root morphology analysis [13,16,48,55], and retreatment prediction [47,60], leaving considerable room for future research on NSRCT outcome prediction.

7. Conclusions

The findings of this study suggest that image-based artificial intelligence models (DL) exhibit superior predictive capability compared to those relying solely on categorical data. Significant improvements in DL were observed compared to professional prognosis (DP), whereas differences among models utilizing categorical data were minimal or statistically insignificant. This finding supports the hypothesis that the information contained in images provides greater richness and discriminatory power in predicting endodontic treatment success compared to categorical data.

These results reinforce the importance of radiographic analysis in evaluating AP and its potential progression, highlighting the critical role of AI models in optimizing clinical diagnoses and therapeutic decision-making. Additionally, further exploration of hybrid models that integrate categorical and imaging data is recommended to enhance predictive accuracy in endodontics.

8. Limitations

Despite the promising findings, this study presents certain limitations that must be considered when interpreting the results. First, the model was developed and validated using a restricted dataset collected from a single institution and obtained using a single radiographic device. This lack of heterogeneity in the sample may affect the generalizability of the results to other populations and clinical settings.

Furthermore, the scarcity of previous studies addressing the prediction of the success of NSRCTs for apical periodontitis using artificial intelligence poses a challenge for comparing and validating our findings against existing literature. The limited availability of specific bibliographic material hinders the direct comparison of our results with other predictive models in endodontics, highlighting the need for further research in this area.

Therefore, we recommend conducting multicenter studies with larger sample sizes and diverse radiographic equipment, as well as integrating complementary clinical data to enhance the applicability of these models in dental practice.

Author Contributions: Conceptualization and methodology, Catalina Bennasar and Antonio Nadal; formal analysis, investigation, data curation, writing—original draft preparation, Catalina Bennasar, Antonio Nadal, Sebastiana Arroyo and Ángel Arturo López-González; writing—review and editing, supervision, Catalina Bennasar, Antonio Nadal, Sebastiana Arroyo, Yolanda Gonzalez-Cid, Ángel Arturo López-González, Francesc Pérez y Pedro Juan Tárraga. All authors have read and agreed to the published version of the manuscript.

Funding: No.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Balearic Islands Research Ethics Committee (IB4015/19IP).

Informed Consent Statement: Patient consent was waived due to the lack of the possibility of identifying participating patients in the datasets.

Data Availability Statement: Full datasets and R scripts are available upon reasonable request to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017 Dec;42:60–88.
2. Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. *Mol Syst Biol.* 2016 Jul 29;12(7).
3. Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical Image Analysis using Convolutional Neural Networks: A Review. *J Med Syst.* 2018 Nov 8;42(11):226.
4. Lin PL, Huang PY, Huang PW. Automatic methods for alveolar bone loss degree measurement in periodontitis periapical radiographs. *Comput Methods Programs Biomed.* 2017 Sep;148:1–11.
5. Ekert T, Krois J, Meinhold L, Elhennawy K, Emara R, Golla T, et al. Deep Learning for the Radiographic Detection of Apical Lesions. *J Endod.* 2019 Jul 1;45(7):917–922.e5.
6. Wang X, Cai B, Cao Y, Zhou C, Yang L, Liu R, et al. Objective method for evaluating orthodontic treatment from the lay perspective: An eye-tracking study. *American Journal of Orthodontics and Dentofacial Orthopedics.* 2016 Oct;150(4):601–10.
7. Alarifi A, AlZubi AA. Memetic Search Optimization Along with Genetic Scale Recurrent Neural Network for Predictive Rate of Implant Treatment. *J Med Syst.* 2018 Nov 17;42(11):202.
8. Sudeep P, Gehlot P, Murali B, Mariswamy A. Artificial intelligence in endodontics: A narrative review. *Journal of International Oral Health.* 2023;15(2):134.
9. Aminoshariae A, Kulild J, Nagendrababu V. Artificial Intelligence in Endodontics: Current Applications and Future Directions. Vol. 47, *Journal of Endodontics.* Elsevier Inc.; 2021. p. 1352–7.
10. Ourang SA, Sohrabniya F, Mohammad-Rahimi H, Dianat O, Aminoshariae A, Nagendrababu V, et al. Artificial intelligence in endodontics: Fundamental principles, workflow, and tasks. *Int Endod J.* 2024 Nov 26;57(11):1546–65.
11. Khanagar SB, Alfadley A, Alfouzan K, Awawdeh M, Alaqla A, Jamleh A. Developments and Performance of Artificial Intelligence Models Designed for Application in Endodontics: A Systematic Review. *Diagnostics.* 2023 Jan 23;13(3):414.
12. Asiri AF, Altuwalah AS. The role of neural artificial intelligence for diagnosis and treatment planning in endodontics: A qualitative review. Vol. 34, *Saudi Dental Journal.* Elsevier B.V.; 2022. p. 270–81.
13. Karobari MI, Adil AH, Basheer SN, Murugesan S, Savadamoorthi KS, Mustafa M, et al. Evaluation of the Diagnostic and Prognostic Accuracy of Artificial Intelligence in Endodontic Dentistry: A Comprehensive

- Review of Literature. Vol. 2023, Computational and mathematical methods in medicine. NLM (Medline); 2023. p. 7049360.
14. Machoy M, Szyszka-Sommerfeld L, Vegh A, Gedrange T, Woźniak K. The ways of using machine learning in dentistry. *Advances in Clinical and Experimental Medicine*. 2020 Mar 24;29(3):375–84.
 15. Zhang Y, Weng Y, Lund J. Applications of Explainable Artificial Intelligence in Diagnosis and Surgery. *Diagnostics*. 2022 Jan 19;12(2):237.
 16. Pethani F. Promises and perils of artificial intelligence in dentistry. Vol. 66, *Australian Dental Journal*. Blackwell Publishing; 2021. p. 124–35.
 17. Schwendicke F, Samek W, Krois J. Artificial Intelligence in Dentistry: Chances and Challenges. *J Dent Res*. 2020 Jul 1;99(7):769–74.
 18. Azarpazhooh A, Khazaei S, Jafarzadeh H, Malkhassian G, Sgro A, Elbarbary M, et al. A Scoping Review of Four Decades of Outcomes in Nonsurgical Root Canal Treatment, Nonsurgical Retreatment, and Apexification Studies: Part 3—A Proposed Framework for Standardized Data Collection and Reporting of Endodontic Outcome Studies. *J Endod*. 2022 Jan;48(1):40–54.
 19. Azarpazhooh A, Sgro A, Cardoso E, Elbarbary M, Laghapour Lighvan N, Badewy R, et al. A Scoping Review of 4 Decades of Outcomes in Nonsurgical Root Canal Treatment, Nonsurgical Retreatment, and Apexification Studies—Part 2: Outcome Measures. Vol. 48, *Journal of Endodontics*. Elsevier Inc.; 2022. p. 29–39.
 20. Azarpazhooh A, Cardoso E, Sgro A, Elbarbary M, Laghapour Lighvan N, Badewy R, et al. A Scoping Review of 4 Decades of Outcomes in Nonsurgical Root Canal Treatment, Nonsurgical Retreatment, and Apexification Studies—Part 1: Process and General Results. Vol. 48, *Journal of Endodontics*. Elsevier Inc.; 2022. p. 15–28.
 21. Bennisar C, García I, Gonzalez-Cid Y, Pérez F, Jiménez J. Second Opinion for Non-Surgical Root Canal Treatment Prognosis Using Machine Learning Models. *Diagnostics*. 2023 Sep 1;13(17).
 22. Lee J, Seo H, Choi YJ, Lee C, Kim S, Lee YS, et al. An Endodontic Forecasting Model Based on the Analysis of Preoperative Dental Radiographs: A Pilot Study on an Endodontic Predictive Deep Neural Network. *J Endod*. 2023 Jun 1;
 23. Li S, Liu J, Zhou Z, Zhou Z, Wu X, Li Y, et al. Artificial intelligence for caries and periapical periodontitis detection. *J Dent*. 2022 Jul 1;122.
 24. Panyarak W, Wantanajittikul K, Suttapak W, Charuakkra A, Prapayastok S. Feasibility of deep learning for dental caries classification in bitewing radiographs based on the ICCMS™ radiographic scoring system. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2023 Feb 1;135(2):272–81.
 25. Mohammad-Rahimi H, Dianat O, Abbasi R, Zahedrozegar S, Ashkan A, Motamedian SR, et al. Artificial Intelligence for Detection of External Cervical Resorption Using Label-Efficient Self-Supervised Learning Method. *J Endod*. 2024 Feb;50(2):144–153.e2.
 26. Kim BS, Yeom HG, Lee JH, Shin WS, Yun JP, Jeong SH, et al. Deep Learning-Based Prediction of Paresthesia after Third Molar Extraction: A Preliminary Study. *Diagnostics*. 2021 Aug 30;11(9):1572.
 27. Vilkomir K, Phen C, Baldwin F, Cole J, Herndon N, Zhang W. Classification of mandibular molar furcation involvement in periapical radiographs by deep learning. *Imaging Sci Dent*. 2024;54(3):257.
 28. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. 2nd ed. New York: Springer New York; 2021. 229–232 p.
 29. El Adoui M, Drisis S, Benjelloun M. Multi-input deep learning architecture for predicting breast tumor response to chemotherapy using quantitative MR images. *Int J Comput Assist Radiol Surg*. 2020 Sep 16;15(9):1491–500.
 30. Mukherjee P, Zhou M, Lee E, Schicht A, Balagurunathan Y, Napel S, et al. A shallow convolutional neural network predicts prognosis of lung cancer patients in multi-institutional computed tomography image datasets. *Nat Mach Intell*. 2020 May 18;2(5):274–82.
 31. Xie H, Zhang T, Song W, Wang S, Zhu H, Zhang R, et al. Super-resolution of Pneumocystis carinii pneumonia CT via self-attention GAN. *Comput Methods Programs Biomed*. 2021 Nov;212:106467.
 32. Ricucci D, Siqueira JF. Biofilms and Apical Periodontitis: Study of Prevalence and Association with Clinical and Histopathologic Findings. *J Endod*. 2010 Aug;36(8):1277–88.

33. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent*. 2018 Oct 1;77:106–11.
34. Kim JE, Nam NE, Shim JS, Jung YH, Cho BH, Hwang JJ. Transfer Learning via Deep Neural Networks for Implant Fixture System Classification Using Periapical Radiographs. *J Clin Med*. 2020 Apr 14;9(4):1117.
35. Krois J, Ekert T, Meinhold L, Golla T, Kharbot B, Wittemeier A, et al. Deep Learning for the Radiographic Detection of Periodontal Bone Loss. *Sci Rep*. 2019 Jun 11;9(1):8495.
36. Hung K, Montalvao C, Tanaka R, Kawai T, Bornstein MM. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: A systematic review. *Dentomaxillofacial Radiology*. 2019;49(1).
37. Yang S, Lee H, Jang B, Kim KD, Kim J, Kim H, et al. Development and Validation of a Visually Explainable Deep Learning Model for Classification of C-shaped Canals of the Mandibular Second Molars in Periapical and Panoramic Dental Radiographs. *J Endod*. 2022 Jul;48(7):914–21.
38. Umer F, Habib S. Critical Analysis of Artificial Intelligence in Endodontics: A Scoping Review. *J Endod*. 2021 Nov;
39. Chugal NM, Clive JM, Spångberg LSW. A prognostic model for assessment of the outcome of endodontic treatment: Effect of biologic and diagnostic variables. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*. 2001 Mar;91(3):342–52.
40. Friedman S. Prognosis of initial endodontic therapy. *Endod Topics*. 2002 Jul 2;2(1):59–88.
41. Jiménez Pinzón A, Segura Egea JJ. Valoración clínica y radiológica del estado periapical: registros e índices periapicales. *Endodoncia (Mex)*. 2003;21(4):220–8.
42. Chugal NM, Clive JM, Spångberg LSW. Endodontic infection: some biologic and treatment factors associated with outcome. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*. 2003 Jul;96(1):81–90.
43. FRIEDMAN S, ABITBOL S, LAWRENCE H. Treatment Outcome in Endodontics: The Toronto Study. Phase 1: Initial Treatment. *J Endod*. 2003 Dec;29(12):787–93.
44. FARZANEH M, ABITBOL S, LAWRENCE H, FRIEDMAN S. Treatment Outcome in Endodontics—The Toronto Study. Phase II: Initial Treatment. *J Endod*. 2004 May;30(5):302–9.
45. Li Y, Zeng G, Zhang Y, Wang J, Jin Q, Sun L, et al. AGMB-Transformer: Anatomy-Guided Multi-Branch Transformer Network for Automated Evaluation of Root Canal Therapy. *IEEE J Biomed Health Inform*. 2022 Apr;26(4):1684–95.
46. Herbst CS, Schwendicke F, Krois J, Herbst SR. Association between patient-, tooth- and treatment-level factors and root canal treatment failure: A retrospective longitudinal and machine learning study. *J Dent [Internet]*. 2022 Feb;117:103937. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0300571221003596>
47. Campo L, Aliaga IJ, De Paz JF, Garcia AE, Bajo J, Villarubia G, et al. Retreatment Predictions in Odontology by means of CBR Systems. *Comput Intell Neurosci*. 2016;2016.
48. Ramezanzade S, Laurentiu T, Bakhshandah A, Ibragimov B, Kvist, Azam, et al. The efficiency of artificial intelligence methods for finding radiographic features in different end. *Acta Odontol Scand [Internet]*. 2023 [cited 2023 Aug 29];81:422–35. Available from: https://uib.gtbib.net:443/menu_usuario.php?p=I0dUYldWMFIXUmhkRzl6TG5Cb2NBPT0=&texto=14303043
49. Bettina Basrani. *Endodontic Radiology*. Segunda. Wiley-Blackwell; 2012. 36–38 p.
50. Alasqah M, Alotaibi FD, Gufran K. The Radiographic Assessment of Furcation Area in Maxillary and Mandibular First Molars while Considering the New Classification of Periodontal Disease. *Healthcare (Switzerland)*. 2022 Aug 1;10(8).
51. Sunnetci KM, Kaba E, Beyazal Çeliker F, Alkan A. Comparative parotid gland segmentation by using <scp>ResNet</scp> -18 and <scp>MobileNetV2</scp> based <scp>DeepLab</scp> v3+ architectures from magnetic resonance images. *Concurr Comput*. 2023 Jan 10;35(1).
52. Kim YH, Park JB, Chang MS, Ryu JJ, Lim WH, Jung SK. Influence of the depth of the convolutional neural networks on an artificial intelligence model for diagnosis of orthognathic surgery. *J Pers Med*. 2021 May 1;11(5).

53. Shan T, Tay FR, Gu L. Application of Artificial Intelligence in Dentistry. Vol. 100, Journal of Dental Research. SAGE Publications Inc.; 2021. p. 232–44.
54. Hwang JJ, Jung YH, Cho BH, Heo MS. An overview of deep learning in the field of dentistry. *Imaging Sci Dent.* 2019 Mar 1;49(1):1–7.
55. Boreak N. Effectiveness of Artificial Intelligence Applications Designed for Endodontic Diagnosis, Decision-making, and Prediction of Prognosis: A Systematic Review. *Journal of Contemporary Dental Practice.* 2020;21(8):926–34.
56. Moidu NP, Sharma S, Chawla A, Kumar V, Logani A. Deep learning for categorization of endodontic lesion based on radiographic periapical index scoring system. *Clin Oral Investig.* 2021;
57. Pauwels R, Brasil DM, Yamasaki MC, Jacobs R, Bosmans H, Freitas DQ, et al. Artificial intelligence for detection of periapical lesions on intraoral radiographs: Comparison between convolutional neural networks and human observers. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2021 May 1;131(5):610–6.
58. Sadr S, Mohammad-Rahimi H, Motamedian SR, Zahedrozegar S, Motie P, Vinayahalingam S, et al. Deep Learning for Detection of Periapical Radiolucent Lesions: A Systematic Review and Meta-analysis of Diagnostic Test Accuracy. *J Endod* [Internet]. 2022 Dec; Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0099239922008457>
59. Khanagar SB, Al-ehaideb A, Maganur PC, Vishwanathaiah S, Patil S, Baeshen HA, et al. Developments, application, and performance of artificial intelligence in dentistry – A systematic review. Vol. 16, Journal of Dental Sciences. Association for Dental Sciences of the Republic of China; 2021. p. 508–22.
60. Sherwood AA, Sherwood AI, Setzer FC, K SD, Shamili J V., John C, et al. A Deep Learning Approach to Segment and Classify C-Shaped Canal Morphologies in Mandibular Second Molars Using Cone-beam Computed Tomography. *J Endod.* 2021 Dec;47(12):1907–16.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.