Digital Dentistry Journal 2 (2025) 100015

Contents lists available at ScienceDirect

Digital Dentistry Journal

journal homepage: www.sciencedirect.com/journal/digital-dentistry-journal

Accuracy and completeness of ChatGPT-4o in the management of non-carious cervical lesions

Ezgi Gurbuz *, Beyza Tetik

Department of Periodontology, Gulsum Gural Faculty of Dentistry, Kutahya Health Sciences University, Kutahya, Turkey

ARTICLE INFO

Keywords:
Artificial intelligence
Case management
Dental caries
Gingival recession
Machine learning

ABSTRACT

Objectives: This study aimed to assess the accuracy and completeness of Chat Generative Pre-trained Transformer (ChatGPT) in controlling noncarious cervical lesions (NCCLs) associated with gingival recession.

Methods: Twelve case scenarios were created using the clinical and radiographic examination data of young adults who came to the periodontology clinic for the first time because of dentin hypersensitivity or esthetic concerns. Gingival recession and NCCL classifications, as well as two reviews, were taken into consideration when developing case scenarios. ChatGPT was asked to offer answers in the fields of diagnosis, clinical management, and surgical management, as well as all bibliographic references used to develop those replies. All replies received during this procedure were examined by four independent reviewers. The reviewers used a 6-point Likert scale to assess the accuracy of each domain response, a 3-point Likert scale for completeness, and the modified global quality scale to assess existing references.

Results: The agreement among the reviewers ranged from 0.629 to 1.000. A statistically significant association was found between the accuracy and completeness ratings in the diagnosis ($p = 0.005$) and clinical management ($p = 0.010$) domains. However, no statistically significant association was discovered between surgical management accuracy and completeness. The majority of the references were of a moderate level, with three showing good quality.

Conclusions: This study found that a full solution that addressed all parts of the cases was rarely possible. At this point, using ChatGPT as a supplemental tool for clinical decision-making in complex NCCL case scenarios does not appear practical.

1. Introduction

The gingival margin is clinically represented by a scalloped line positioned 1–2 mm coronal to the cementoenamel junction (CEJ) [1]. Gingival recession is defined as the apical displacement of the gingival margin relative to the CEJ, resulting in exposure of the root surface to the oral cavity [1]. This condition can lead to esthetic concerns and dentin hypersensitivity [2]. Additionally, it may be associated with various dental conditions, including noncarious cervical lesions (NCCLs) [2].

An NCCL is characterized by the loss of tooth structure at the cervical level, which is not caused by dental caries [3]. The etiology of NCCLs is complex, involving a combination of factors such as biocorrosion (erosion), friction (abrasion), and stress (abfraction) [4].

An NCCL may affect only the crown, the root surface, or both the crown and the exposed root [5]. When the lesion involves the root surface, it is often linked to gingival recession, and modifications to the root surface, such as CEJ loss and concavity formation, are commonly observed [2].

Root coverage procedures are employed to treat gingival recession, with success typically measured by the percentage of root coverage. This is calculated by comparing the preoperative and postoperative recession measurements [6]. The CEJ is a critical reference point for determining the extent of recession [7]. In cases where the CEJ cannot be identified due to the presence of an NCCL, the diagnosis, treatment, and prognosis may be compromised, as the recession extension cannot be accurately assessed. Several methods for identifying the missing CEJ have been proposed to address this issue [6,7]. Zucchelli et al. [6,8] suggested determining the CEJ's location by connecting reference points on the mesial and distal line angles, which are also used to calculate the ideal vertical dimension of the interdental papilla. Cairo and Pini-Pirato [7]

*Corresponding author. Department of Periodontology, Gulsum Gural Faculty of Dentistry, Kutahya Health Sciences University, Inkoy, Eskisehir highway, No:65, Kutahya, 43100, Turkey.

E-mail addresses: ezgi.dogan@ksbu.edu.tr (E. Gurbuz), beyzaa.tetikk@hotmail.com (B. Tetik).

https://doi.org/10.1016/j.ddj.2025.100015
Received 7 February 2025; Received in revised form 29 April 2025; Accepted 30 April 2025
Available online 8 May 2025

recommended measuring the crown length and width of a contralateral homologous tooth or an adjacent tooth without gingival recession to estimate the missing CEJ's position.

Based on the CEJ determination methods, several treatment strategies have been recommended [5,9]. These strategies differ from those used for gingival recessions without surface modifications and require a combination of restorative and periodontal approaches [10]. However, many dentists lack sufficient knowledge regarding the management and treatment of NCCLs [11,12]. Moreover, root coverage procedures and combined restorative and periodontal approaches are not widely considered viable alternatives [11].

Artificial intelligence offers large language models (LLMs) that use extensive training datasets, including internet sources such as books, articles, and websites [13]. Chat Generative Pre-trained Transformer (ChatGPT; OpenAI Inc., San Francisco, California) is an LLM chatbot capable of generating human-like conversational responses to various queries [14]. ChatGPT-4o, the latest and most advanced model, was released in May 2024 [15].

Large language model chatbots have a variety of applications in dental medicine [14]. These chatbots can rapidly assimilate, summarize, and rephrase information, thus reducing the administrative burden on clinicians [16]. By collecting patient information and symptoms, chatbots can also enhance dental telemedicine services, particularly in underserved areas [17]. Furthermore, LLM chatbots provide significant support to students and healthcare professionals in tasks such as text summarization, translation, and education [14,18]. They can also inform patients about dental procedures [19,20]. Additionally, these chatbots have the potential to assist in clinical decision-making [21–23]. However, several issues are impeding clinical deployment [16]. The training datasets are insufficient to provide up-to-date information, and the responses require confirmation for domain-specific accuracy. Furthermore, the outputs are based on learned connections between words rather than an understanding of the information contained in the responses. In this approach, incorrect information, termed "hallucinations," is also possible [24,25]. Given the limitations of LLM chatbots, autonomous deployment in clinical dentistry is presently not practical. Semi-autonomous deployment, including physicians, may be appropriate, and pragmatic studies are required to validate the chatbots successfully [16]. Given that NCCL patients linked with gingival recession are an area in which dentists lack information, ChatGPT may be a useful tool in the management of these cases in ordinary clinical practice. The current study aimed to assess the accuracy and completeness of ChatGPT-4o in treating clinical scenarios of NCCLs with gingival recession.

2. Materials and methods

This study adhered to the principles of the Declaration of Helsinki. Ethics committee approval was waived as no human or animal participants were involved. The study was conducted in accordance with the STROBE guidelines [26].

2.1. Case scenarios

Two periodontists developed 12 case scenarios related to NCCL. Each scenario comprised patient complaints, symptoms, and comprehensive clinical and radiological tests. Case scenarios were presented based on the clinical and radiographic examination results of young adults with no substantial medical history who were visiting the periodontology clinic for the first time. The clinical examination revealed that the reason for referral was dentin hypersensitivity or esthetic problems relating to a single tooth. This was followed by a description of the lesion's morphological features in the labial/buccal area, the depth of the surface discrepancy, the absence of carious lesions, and the CEJ's determinability. Caries and interproximal attachment loss were documented radiographically.

The initial scenario was determined using a classification of gingival recession and a classification of NCCL [27,28]. An evidence-based review presenting a decision-making algorithm for treating NCCLs, along with a detailed review, was considered in developing the text of the case scenarios [9,29]. A detailed review was used to establish the morphological features of the lesion [29]. Pini-Pirato et al. [28] classified dental surface defects in areas of gingival recession. Cairo et al. [27] examined interproximal attachment loss and divided instances into three categories: RT1, RT2, and RT3. The evidence-based approach described by Santamaria et al. [9] enhanced the two classifications above by integrating the step depth, leading to the definition of clinical conditions as RT1A−, RT1A+, RT1A+V, RT1B−, RT1B+, RT1B+V, and RT2A−. To add complexity, case RT2B+V was included, which highlighted the absence of the CEJ and the presence of a profound defect with interproximal attachment loss. In addition, four more cases were included to account for the thin phenotype, severe dentin hypersensitivity, esthetic issue in the presence of CEJ, and esthetic concern in the absence of CEJ, bringing the total to 12. Finally, two NCCL cases with interproximal attachment loss and 10 NCCL cases with no attachment loss were created. The NCCLs without attachment loss were further characterized according to the presence of CEJ or tooth surface inconsistencies. The CEJ could not be established in half of the cases, and seven showed a surface discrepancy. The case scenarios are described in Appendix A, which is available online.

2.2. ChatGPT-4o analysis

On May 27, 2024, a single researcher inputted all case scenarios into ChatGPT-4o. No specifically created agent was used for this study. A detailed prompt was crafted to simulate a dialogue between a periodontist and a general dentist (e.g., "Imagine that you are a periodontist, and I am a general dentist. Please answer the following question accurately and directly, without rambling or providing creative answers") [30]. ChatGPT-4o was asked to respond to three domains: diagnosis, clinical management, and surgical management, based on the details of each case scenario (e.g., "What are your thoughts on the diagnosis? How would you manage the patient clinically and surgically? Provide a detailed diagnosis and, if necessary, nonsurgical and surgical procedures" (Appendix A) [23]. Additionally, ChatGPT-4o was instructed to provide all bibliographic references used in formulating its responses. Each scenario was submitted as a separate chat. After receiving a response for each case, the response was copied into a Microsoft Word document (Redmond, Washington), and the chat was subsequently deleted.

2.3. Evaluation of the ChatGPT-4o answers

All responses were meticulously documented and thoroughly analyzed by four independent reviewers. The reviewers were specialists who had graduated from an accredited dental school and completed a 3-yr postgraduate periodontal training program. They had extensive experience in general dentistry, with at least 5 yr of specialized practice. Before reviewing the responses, the reviewers researched the classifications and evidence-based algorithms used to prepare the cases.

Each response was evaluated independently for accuracy and completeness [31]. A 6-point Likert scale was used to assess accuracy, where 1 represented a completely incorrect response, 2 indicated more incorrect than correct, 3 indicated an equal balance of correct and incorrect, 4 represented more correct than incorrect, 5 indicated nearly all correct, and 6 indicated completely correct. Completeness was assessed using a 3-point Likert scale: 1 represented an incomplete answer with significant parts missing, 2 indicated an adequate answer addressing all aspects of the question, and 3 represented a comprehensive response that covered all aspects of the question and provided additional information exceeding expectations. Each domain (diagnostic, clinical care, and surgical management) was evaluated using these accuracy and completeness scales. The final scores for each domain were calculated by comparing the reviewers' results. In cases where discrepancies were found, consensus was reached.

The reviewers also assessed the sources cited in ChatGPT-4o's responses. The references were initially verified through internet searches using Google, Google Scholar, and PubMed. References were considered valid if any part of the reference (e.g., full author names, complete title, abstract, or full text) could be verified. The quality of each reference was then assessed based on its relevance to the clinical condition, timeliness, and the degree of support it provided to the response. The Modified Global Quality Scale was used to assign a score to each reference [31]. When a reference appeared in multiple case scenarios, it was evaluated within the context of each relevant scenario. Consensus scores from the reviewers were recorded.

2.4. Statistical analysis

Statistical analysis was performed using SPSS for Windows (version 25, IBM Corp., Armonk, New York). Cohen's κ statistic was used to assess the agreement between the categorical measurement values generated by the reviewers. The Shapiro–Wilk test was applied to validate the assumption of normality for data distribution. The Pearson correlation test was used to examine the linear relationship between continuous variables with a normal distribution. In cases where the sample size assumption (expected value > 5) for analyzing the association between categorical variables was not met, the Fisher exact test was used. p values < 0.05 were considered statistically significant.

3. Results

The κ statistics estimated between the reviewers for the research domains ranged from 0.629 to 1.000 (Table 1). The interrater reliability was shown to be statistically significant and consistently high, with a minimum reliability level >0.610.

Table 2 shows the distributions of accuracy and completeness measurements and the relationships between these values throughout the three domains. In response to the five case scenarios, the diagnosis received an accuracy score of 4 (more correct than incorrect). Half of the responses indicated that the diagnosis was nearly all correct, with only one NCCL instance receiving an entirely correct score. When the distribution of clinical and surgical management accuracy ratings was examined, the majority of cases ($n = 8$, 66.7 %) earned a score of 4 in both domains. In clinical management, one case was graded as completely accurate, whereas in surgical management, none were scored as entirely correct. In terms of completeness, the three domains produced distinct results: two instances in the diagnosis domain obtained a score of 3, while no cases in the clinical management area had a thorough response. In the surgical management domain, one case received a score of 3.

A statistically significant relationship was found between the accuracy and completeness scores in the diagnosis ($p = 0.005$) and clinical management ($p = 0.010$) domains (Table 2). Cases with a diagnosis accuracy score of 5 often had a completeness score of 2. Cases having a clinical management accuracy score of 4 often had a completeness score of 1. No statistically significant connection was found between surgical management accuracy and completeness. Similarly, no significant linear correlation was detected between the total accuracy and completeness scores. (Table 3).

Table 1
Kappa statistics among reviewers.

Domains	Kappa Statistics	p
Diagnosis Accuracy	0.629	0.001*
Clinical Management Accuracy	1.000	<0.001*
Surgical Management Accuracy	0.707	0.001*
Diagnosis Completeness	1.000	<0.001*
Clinical Completeness	0.700	0.013*
Surgical Completeness	0.852	<0.001*

*$p < 0.05$.

There was no association discovered between the distributions of accuracy scores in the three domains, as well as completeness scores (Tables 4 and 5).

ChatGPT-4o recommended 92 references for the 12 NCCL case scenarios. It was revealed that 39 (42.4 %) of the references were nonexistent. After a thorough inspection of the remaining 53 references, it was discovered that 23 (25 %) had mistakes in terms of authors, year, volume, issue, page numbers, digital object identification, or PubMed identifiers. The majority of the references (39.6 %) had intermediate quality, with three (5.7 %) having good quality (Table A.1). None of the recommended references demonstrated high quality.

4. Discussion

The decision-making process for the treatment of NCCL is a complex interdisciplinary issue [5,9]. Both general dentists and specialist practitioners lack appropriate knowledge [11,12]. Nowadays, the application of new technologies can help clinicians make better decisions. This study assessed the accuracy and completeness of ChatGPT-4o during the decision-making process for NCCL treatment. Throughout the procedure, ChatGPT-4o was responsible for reviewing case scenarios involving diagnosis, clinical care, and surgical management. Despite delivering proper answers in all three categories, a full response that addressed all facets of the scenario was unusual. NCCLs were frequently referenced in diagnostic suggestions; however, the relationship between NCCL and gingival recession was only mentioned in two cases. In clinical management, precise recommendations and detailed suggestions for addressing the treatment of abrasion, erosion, and abfraction were rarely provided. Restorative treatment was only appropriately addressed in a few cases. When an NCCL is associated with gingival recession, the decision to apply restorative treatment depends on factors such as the severity of dentin hypersensitivity, the localization of the NCCL to the crown, the depth of the defect, and the presence of interproximal attachment loss [9]. While information regarding dentin hypersensitivity, defect depth, and attachment loss was provided in the presented cases, ChatGPT's requirement to infer whether the lesion involved the crown or both the crown and root, based on CEJ determination, may have made it difficult to accurately recommend restorative treatment.

From a surgical management perspective, comprehensive responses addressing specific issues, such as patient complaints, keratinized tissue thickness, defect depth, interproximal attachment level, and CEJ determination, were infrequent. Overall, combined restorative and periodontal treatment was rarely emphasized. Additionally, the relationship between restoration and CEJ was not discussed at all.

In a comparable study on odontogenic rhinosinusitis case scenarios, ChatGPT's performance was found to be suboptimal [23]. It was noted that ChatGPT failed to identify certain nuances in the cases and did not propose treatment plans based on these nuances. Similarly, as observed in this study, ChatGPT displayed a wide range of predictions and flexibility in its responses, rather than generating a specific treatment plan.

In a multicenter analysis assessing the accuracy of information generated by ChatGPT in the context of head and neck and oromaxillofacial surgery, it was determined that responses to clinical scenarios were 81.7 % accurate and 56.7 % complete [31]. Compared with the results of this study, overall diagnostic and therapeutic outcomes were higher. A study on oral pathology case scenarios revealed that the responses provided had a high agreement with the evaluators' diagnoses and treatment strategies [21]. Upon reviewing the literature, it was noted that most studies involving ChatGPT in dentistry are focused on oral surgery [19,21,23,30–32]. This could result in enhanced efficiency for ChatGPT in well-studied areas while delivering reduced accuracy and completeness rates in freshly explored fields, as demonstrated in the present study. This could be attributable to the iterative nature of ChatGPT's training process and the reinforcement learning gained from human feedback [33]. To date, no study has focused on case scenarios developed in the field of periodontology. In a study addressing a similarly

Table 2
Relationship between accuracy and completeness measurements in three domains.

Completeness		1			2			3				
Accuracy	Score	n	%	C %	n	%	C %	n	%	C %	Test Statistics	*p*
Diagnosis	4	3	60.0	100.0	1	20.0	14.3	1	20.0	50.0	10.213	0.005*
	5	0	0.0	0.0	6	100.0	85.7	0	0.0	0.0		
	6	0	0.0	0.0	0	0.0	0.0	1	100.0	50.0		
Clinical Management	3	0	0.0	0.0	1	100.0					7.632	0.010*
	4	7	87.5	100.0	1	12.5						
	5	0	0.0	0.0	2	100.0						
	6	0	0.0	0.0	1	100.0						
Surgical Management	3	0	0.0	0.0	1	100.0	25.0	0	0.0	0.0	5.827	0.253
	4	6	75.0	85.7	1	12.5	25.0	1	12.5	100.0		
	5	1	33.3	14.3	2	66.7	50.0	0	0.0	0.0		

*$p < 0.05$, %: row percentage and C %: column percentage for completeness.

Table 3
Linear correlation between total accuracy and completeness scores.

	Min-Max	Mean ± SD (Median)	r	*p*
Accuracy	11–15	13.08 ± 1.16 (13)	0.454	0.138
Completeness	3–7	4.83 ± 1.40 (5)		

Min: minimum; max: maximum; SD: standard deviation.

unusual topic and incorporating case scenarios relating to interceptive orthodontics, it was discovered that ChatGPT-4 did not yield responses that were 100 % correct and complete [22]. Although similar results have been achieved in rarely studied topics, ChatGPT is a current issue, and there is no evidence on the effect of repeating the same topic in different trials on ChatGPT proficiency.

In this study, it was discovered that 42.4 % of the references provided by ChatGPT-4o were missing. Vaira et al. [31] found a comparable rate in their ChatGPT study on head and neck surgery. However, research conducted on scientific writing produced a lower incidence (16 %) [24].

Although studies have demonstrated acceptable accuracy rates, the consensus is that ChatGPT is still far from being a reliable tool for decision-making support [19,21,22,31]. This is primarily due to several limitations inherent in ChatGPT. One such limitation is the extensive dataset on which ChatGPT is trained. In this study, it was observed that the references used in case preparation were not among those ChatGPT relied upon in its responses. The quality of the references, as indicated by the reference quality score in the present study, suggests that the

Table 4
Relationship between accuracy measurements in three domains.

Diagnosis	Score	4			5			6			Test Statistics	*p*
		n	%	D %	n	%	D %	n	%	D %		
Clinical Management	3	0	0.0	0.0	1	100.0	16.7	0	0.0	0.0	5.143	1.000
	4	4	50.0	80.0	3	37.5	50.0	1	12.5	100.0		
	5	1	50.0	20.0	1	50.0	16.7	0	0.0	0.0		
	6	0	0.0	0.0	1	100.0	16.7	0	0.0	0.0		
Surgical Management	3	0	0.0	0.0	1	100.0	16.7	0	0.0	0.0	3.143	1.000
	4	4	50.0	80.0	3	37.5	50.0	1	12.5	100.0		
	5	1	33.3	20.0	2	66.7	33.3	0	0.0	0.0		

Clinical Management	Score	3			4			5			6			Test Statistics	*p*
		n	%	CM %	n	%	CM %	n	%	CM %	n	%	CM %		
Surgical Management	3	1	100.0	100.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0	8.358	0.352
	4	0	0.0	0.0	5	62.5	62.5	2	25.0	100.0	1	12.5	100.0		
	5	0	0.0	0.0	3	100.0	37.5	0	0.0	0.0	0	0.0	0.0		

%: row percentage, D %: column percentage for diagnosis and CM %: column percentage for clinical management.

Table 5
Relationship between completeness measurements in three domains.

Diagnosis	Score	1			2			3			Test Statistics	*p*
		n	%	D %	n	%	D %	n	%	D %		
Clinical Management	1	3	42.9	100.0	3	42.9	42.9	1	14.3	50.0	2.753	0.308
	2	0	0.0	0.0	4	80.0	57.1	1	20.0	50.0		
Surgical Management	1	3	42.9	100.0	3	42.9	42.9	1	14.3	50.0	3.686	0.660
	2	0	0.0	0.0	3	75.0	42.9	1	25.0	50.0		
	3	0	0.0	0.0	1	100.0	14.3	0	0.0	0.0		

Clinical Management	Score	1			2			Test Statistics	*p*
		n	%	CM %	n	%	CM %		
Surgical Management	1	5	71.4	71.4	2	28.6	40.0	1.982	0.376
	2	2	50.0	28.6	2	50.0	40.0		
	3	0	0.0	0.0	1	100.0	20.0		

%: row percentage, D %: column percentage for diagnosis and CM %: column percentage for clinical management.

unsupervised training data was not relevant to the topic. This limitation could be addressed by training ChatGPT on context-specific, smaller datasets, supplemented with supervised learning techniques [34]. Additionally, LLM chatbots can be paired with retrieval-augmented generation models to provide a database of verified information [35]. This approach may reduce hallucinations and facilitate access to up-to-date references. Future studies should assess the impact of these models on the accuracy of information and the quality of references.

In addition to ChatGPT, other artificial intelligence–powered conversational agents include Google Gemini (formerly known as Bard; Mountain View, California) and Microsoft Copilot (formerly known as Bing; Redmond, Washington). A study comparing the accuracy of ChatGPT-4, ChatGPT-3.5, Bing, and Bard in 11 fictional endodontic cases found that ChatGPT-4 led the group in providing diagnostic and treatment recommendations [36]. Another study, where four artificial intelligence models—ChatGPT-4, Google Gemini, Google Gemini Advanced, and Microsoft Copilot—were tasked with answering 10 open-ended questions about periodontal disease treatment, reported that ChatGPT-4 outperformed the other models [37]. However, differing results have been observed across various fields, including orthodontics and oral surgery [38,39]. It is apparent that the models offered in the ever-changing digital world must be compared to multiple research studies completed in other fields.

The study had some drawbacks. First, difficulties were found while executing the accuracy grading in the clinical and surgical management domains. Although no incorrect information was presented, a "nearly all correct" grade was given because ChatGPT indicated treatment planning in an irregular manner rather than the recommended order in the reference article [9]. Unfortunately, there is no established rating system for LLM chatbots. Second, a small number of cases were examined. The examples were produced utilizing a method that took into account the whole range of clinical manifestations and the literature's classifications and decision-making processes. Third, despite the postulated relationship between NCCL's morphological qualities and the etiological variables of erosion, abrasion, and abfraction, the case scenarios did not specifically address these etiological issues. Instead, it was expected that ChatGPT would deduce the etiology based on morphological characteristics. Although the case report mentioned dentin hypersensitivity and a cervical defect without cavities, it did not go into depth on a full periodontal examination or stimulation, vitality, percussion, and palpation testing. Finally, considering the quick improvements in digital technology and the great improvement of LLMs' reasoning, it would be useful to evaluate the performance of different LLMs in these complex scenarios.

5. Conclusions

This context-specific study found that ChatGPT-4o delivered the majority of the right answers in the diagnosis, clinical management, and surgical management domains of NCCL cases. However, obtaining a full response that addressed all facets of the scenario was unusual. Furthermore, ChatGPT-4o did not provide references of high quality. At this point, using ChatGPT as a supplemental tool for clinical decision-making in complex NCCL case scenarios does not appear practical. Further research is required to examine clinical circumstances in many areas, where LLM chatbots—whether integrated with retrieval-based models or not—are compared with other chatbots to increase accuracy and reference quality.

CRediT authorship contribution statement

Ezgi Gurbuz: Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Beyza Tetik:** Writing – original draft, Methodology, Data curation, Conceptualization.

Funding

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

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ddj.2025.100015.

Data availability

Data will be made available on request.

References

[1] Mucogingival surgery, in: N.P. Lang, T. Karring (Eds.), Proceedings of the 1st European Workshop on Periodontology, Quintessence, 1994, pp. 193–209.
[2] P. Cortellini, N.F. Bissada, Mucogingival conditions in the natural dentition: narrative review, case definitions, and diagnostic considerations, J. Periodontol. 89 (Suppl 1) (2018) S204–S213, https://doi.org/10.1002/JPER.16-0671.
[3] T.C. Aw, X. Lepe, G.H. Johnson, et al., Characteristics of noncarious cervical lesions: a clinical investigation, J. Am. Dent. Assoc. 133 (6) (2002) 725–733, https://doi.org/10.14219/jada.archive.2002.0268.
[4] M. Peumans, J. De Munck, A. Mine, et al., Clinical effectiveness of contemporary adhesives for the restoration of non-carious cervical lesions. A systematic review, Dent. Mater. 30 (10) (2014) 1089–1103, https://doi.org/10.1016/j.dental.2014.07.007.
[5] G. Zucchelli, G. Gori, M. Mele, et al., Non-carious cervical lesions associated with gingival recessions: a decision-making process, J. Periodontol. 82 (12) (2011) 1713–1724, https://doi.org/10.1902/jop.2011.110080.
[6] G. Zucchelli, T. Testori, M. De Sanctis, Clinical and anatomical factors limiting treatment outcomes of gingival recession: a new method to predetermine the line of root coverage, J. Periodontol. 77 (4) (2006) 714–721, https://doi.org/10.1902/jop.2006.050038.
[7] F. Cairo, G.P. Pini-Prato, A technique to identify and reconstruct the cementoenamel junction level using combined periodontal and restorative treatment of gingival recession. A prospective clinical study, Int. J. Periodontics Restor. Dent. 30 (6) (2010) 573–581.
[8] G. Zucchelli, M. Mele, M. Stefanini, et al., Predetermination of root coverage, J. Periodontol. 81 (7) (2010) 1019–1026, https://doi.org/10.1902/jop.2010.090701.
[9] M.P. Santamaria, I.F. Mathias-Santamaria, L.F.F. Ferraz, et al., Rethinking the decision-making process to treat gingival recession associated with non-carious cervical lesions, Braz. Oral Res. 35 (Supp 2) (2021) e096, https://doi.org/10.1590/1807-3107bor-2021.vol35.0096.
[10] C.A. Silveira, I.F. Mathias, F.L. da Silva Neves, et al., Connective tissue graft and crown-resin composite restoration for the treatment of gingival recession associated with noncarious cervical lesions: case series, Int. J. Periodontics Restor. Dent. 37 (4) (2017) 601–607, https://doi.org/10.11607/prd.3044.
[11] J.C.M. Lee, M.F. Burrow, M.G. Botelho, A qualitative analysis of dentists' understanding and management of non-carious cervical lesions (NCCL), J. Dent. 136 (2023) 104640, https://doi.org/10.1016/j.jdent.2023.104640.
[12] R.A. Modena, P.N. Tannure, V.A. Pessoa, et al., Diagnosis attitudes and restorative practices of non-carious cervical lesions by a group of Brazilian dentists from the state of Rio de Janeiro, Appl Adhes Sci 6 (2018) 1–9, https://doi.org/10.1186/s40563-018-0112-5.
[13] A. Radford, J. Wu, R. Child, et al., Language models are unsupervised multitask learners, OpenAI blog 1 (8) (2019) 9.
[14] F. Eggmann, R. Weiger, N.U. Zitzmann, et al., Implications of large language models such as ChatGPT for dental medicine, J. Esthetic Restor. Dent. 35 (7) (2023) 1098–1102, https://doi.org/10.1111/jerd.13046.
[15] Z. Feng, B. Li, F. Liu, A First Look at Financial Data Analysis Using ChatGPT-4o, Available at: SSRN 4849578, 2024.
[16] A.J. Thirunavukarasu, D.S.J. Ting, K. Elangovan, et al., Large language models in medicine, Nat. Med. 29 (8) (2023) 1930–1940, https://doi.org/10.1038/s41591-023-02448-8.
[17] Lahat A, Klang E. Can advanced technologies help address the global increase in demand for specialized medical care and improve telehealth services? J. Telemed. Telecare Published online February 9, 2023. doi:10.1177/1357633X231155520.

[18] Danesh A, Pazouki H, Danesh F, et al. Artificial intelligence in dental education: ChatGPT's performance on the periodontic in-service examination. J. Periodontol. Published online January 10, 2024. doi:10.1002/JPER.23-0514.
[19] Y. Balel, Can ChatGPT be used in oral and maxillofacial surgery? J. Stomatol. Oral Maxillofac. Surg. 124 (5) (2023) 101471 https://doi.org/10.1016/j.jormas.2023.101471.
[20] O. Babayiğit, Z. Tastan Eroglu, D. Ozkan Sen, et al., Potential use of ChatGPT for patient information in periodontology: a descriptive pilot study, Cureus 15 (11) (2023) e48518, https://doi.org/10.7759/cureus.48518.
[21] Ö. Uranbey, F. Özbey, Ö. Kaygısız, et al., Assessing ChatGPT's diagnostic accuracy and therapeutic strategies in oral pathologies: a cross-sectional study, Cureus 16 (4) (2024) e58607, https://doi.org/10.7759/cureus.58607.
[22] A. Hatia, T. Doldo, S. Parrini, et al., Accuracy and completeness of ChatGPT-generated information on interceptive orthodontics: a multicenter collaborative study, J. Clin. Med. 13 (3) (2024) 735, https://doi.org/10.3390/jcm13030735.
[23] A.M. Saibene, F. Allevi, C. Calvo-Henriquez, et al., Reliability of large language models in managing odontogenic sinusitis clinical scenarios: a preliminary multidisciplinary evaluation, Eur. Arch. Otorhinolaryngol. 281 (4) (2024) 1835–1841, https://doi.org/10.1007/s00405-023-08372-4.
[24] S.A. Athaluri, S.V. Manthena, V. Kesapragada, et al., Exploring the boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references, Cureus 15 (4) (2023) e37432, https://doi.org/10.7759/cureus.37432.
[25] A.J. Thirunavukarasu, R. Hassan, S. Mahmood, et al., Trialling a large language model (ChatGPT) in general practice with the applied knowledge test: observational study demonstrating opportunities and limitations in primary care, JMIR Med Educ 9 (1) (2023) e46599, https://doi.org/10.2196/46599.
[26] E. von Elm, D.G. Altman, M. Egger, et al., The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies, J. Clin. Epidemiol. 61 (4) (2008) 344–349, https://doi.org/10.1016/j.jclinepi.2007.11.008.
[27] F. Cairo, M. Nieri, S. Cincinelli, et al., The interproximal clinical attachment level to classify gingival recessions and predict root coverage outcomes: an explorative and reliability study, J. Clin. Periodontol. 38 (7) (2011) 661–666, https://doi.org/10.1111/j.1600-051X.2011.01732.x.
[28] G. Pini-Prato, D. Franceschi, F. Cairo, et al., Classification of dental surface defects in areas of gingival recession, J. Periodontol. 81 (6) (2010) 885–890, https://doi.org/10.1902/jop.2010.090631.
[29] M. Peumans, G. Politano, B. Van Meerbeek, Treatment of noncarious cervical lesions: when, why, and how, Int J Esthet Dent 15 (1) (2020) 16–42.
[30] A. Suárez, J. Jiménez, M. Llorente de Pedro, et al., Beyond the Scalpel: assessing ChatGPT's potential as an auxiliary intelligent virtual assistant in oral surgery, Comput. Struct. Biotechnol. J. 24 (2024) 46–52, https://doi.org/10.1016/j.csbj.2023.11.058.
[31] L.A. Vaira, J.R. Lechien, V. Abbate, et al., Accuracy of ChatGPT-generated information on head and neck and oromaxillofacial surgery: a multicenter collaborative analysis, Otolaryngol. Head Neck Surg. 170 (6) (2024) 1492–1503, https://doi.org/10.1002/ohn.489.
[32] A.H. Acar, Can natural language processing serve as a consultant in oral surgery? J. Stomatol. Oral Maxillofac. Surg. 125 (3) (2024) 101724 https://doi.org/10.1016/j.jormas.2023.101724.
[33] D. Kalla, N. Smith, F. Samaah, et al., Study and analysis of chat GPT and its impact on different fields of study, Int. J. Innovat. Sci. Res. Technol. 8 (3) (2023) 827–833.
[34] C. Chakraborty, S. Pal, M. Bhattacharya, et al., Overview of Chatbots with special emphasis on artificial intelligence-enabled ChatGPT in medical science, Front. Artif. Intell. 6 (2023) 1237704, https://doi.org/10.3389/frai.2023.1237704.
[35] F. Fanelli, M. Saleh, P. Santamaria, et al., Development and comparative evaluation of a reinstructed GPT-4o model specialized in periodontology, J. Clin. Periodontol. 26 (2024), https://doi.org/10.1111/jcpe.14101. Published online December.
[36] J.D. Mendonça de Moura, C.E. Fontana, V.H. Reis da Silva Lima, et al., Comparative accuracy of artificial intelligence chatbots in pulpal and periradicular diagnosis: a cross-sectional study, Comput. Biol. Med. 183 (2024) 109332, https://doi.org/10.1016/j.compbiomed.2024.109332.
[37] G.S. Chatzopoulos, V.P. Koidou, L. Tsalikis, et al., Large language models in periodontology: assessing their performance in clinically relevant questions, J. Prosthet. Dent 18 (2024), https://doi.org/10.1016/j.prosdent.2024.10.020. Published online November.
[38] A. Azadi, F. Gorjinejad, H. Mohammad-Rahimi, et al., Evaluation of AI-generated responses by different artificial intelligence chatbots to the clinical decision-making case-based questions in oral and maxillofacial surgery, Oral Surg. Oral Med. Oral Pathol. Oral Radiol. 137 (6) (2024) 587–593, https://doi.org/10.1016/j.oooo.2024.02.018.
[39] M.A. Makrygiannakis, K. Giannakopoulos, E.G. Kaklamanos, Evidence-based potential of generative artificial intelligence large language models in orthodontics: a comparative study of ChatGPT, Google Bard, and Microsoft Bing, Eur. J. Orthod. (2024), https://doi.org/10.1093/ejo/cjae017. Published online April 13.