Scholars Journal of Dental Sciences
Abbreviated Key Title: Sch J Dent Sci
ISSN 2394-4951 (Print) | ISSN 2394-496X (Online)
Journal homepage: https://saspublishers.com

OPEN ACCESS

# The Impact of Artificial Intelligence on Contemporary Orthodontic Treatment Planning - A Systematic Review and Meta-Analysis

M'hamed Jihed[1*], Ines Dallel[2], Samir Tobji[2], Adel Ben Amor[2,3]

[1]Researcher, Doctor (DMD), University of Monastir, Faculty of Dental Medicine, 5000 Monastir, Tunisia;
[2]Professors, PHD, University of Monastir, Faculty of Dental Medicine, Dento-Facial Orthopedics Department of Monastir Dental Clinic, Laboratory of Oral Health and Orofacial Rehabilitation, LR12ES11, 5000 Monastir, Tunisia
[3]Head of Dento-Facial Orthopedics Department of Monastir Dental Clinic

**DOI:** 10.36347/sjds.2022.v09i05.001 | **Received:** 25.04.2022 | **Accepted:** 30.05.2022 | **Published:** 04.06.2022

***Corresponding author**: Dr. M'hamed Jihed
Researcher, Doctor (DMD), University of Monastir, Faculty of Dental Medicine, 5000 Monastir, Tunisia

**Abstract** **Original Research Article**

***Objective:*** To assess available evidence on the use of artificial intelligence (AI) in the planning of customized orthodontic therapy. The aim of the meta-analysis was to evaluate the performance and effectiveness of AI models for orthodontic treatment planning and decision-making. ***Materials and methods:*** PubMed, EBSCO host, ScienceDirect, Scopus, and Web of Science were searched over the period from January 1, 2000 to January 9, 2021, then they were updated until January 19, 2022. A systematic review and diagnostic test accuracy meta-analysis were performed. ***Results:*** Overall, 1037 records were identified. A total of twelve studies were ultimately included in the qualitative synthesis, of which five studies were included in the meta-analysis. Pooled sensitivity, specificity, diagnostic odds ratio, and area under the curve with 95% confidence intervals of AI models' performance were: 0.965 (0.921-0.985), 0.962 (0878-0.989), 695.537 (232.742-2078.572), 0.99 (0.98-1.00), respectively. The accuracy of AI systems reached 95.47%. ***Conclusions:*** The findings show promising results concerning the diagnostic accuracy of AI systems for orthodontic treatment planning and decision-making and its implementation in clinical settings. AI models are successful in predicting valid treatment plans with accurate decisions. Thus, they can ease global treatment and improve outcomes.

**Keywords:** Artificial intelligence, technology, orthodontics, treatment-planning, systematic review, meta-analysis.



## INTRODUCTION

In the 21st century, we are witnessing rapid progress in computer technologies and data science along with their potential applications in orthodontics. Consequently, these advancements and emerging technologies affected healthcare and orthodontic research by introducing countless possibilities of developing precise solutions that can ease therapeutic care and enhance outcomes [1].

One of those technologies is artificial intelligence (AI) which basically consists of the development of computer systems that perform tasks usually requiring human intelligence [2]. This technology can help clinicians with the decision-making process, thus, saving time and resources while boosting the treatment's efficiencies.

AI is considered one of the main interests in the scientific community of this decade [3]. To the best of our knowledge, there has been no research that quantitatively assessed the performance of AI systems in planning orthodontic treatments. Hence, the need for a review that gathers all available evidence on the clinical use of this emerging technology, determining its impact in practice and evaluating its effectiveness.

This review was conducted to investigate the impact of AI on contemporary orthodontics through the following question: "What is the effectiveness and performance of artificial intelligence, in orthodontic treatment planning and decision-making compared to reference standards?".

## MATERIALS AND METHODS

### Protocol and Registration

The protocol has been registered since 27th of February 2021 in PROSPERO (CRD42021230816). This review was designed and reported conforming to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 [4] and adhered to the Cochrane guidelines [5]. Approval for conducting this research was granted by the "Thesis committee" of the Faculty of Dental Medicine of Monastir in July 2021.

### Eligibility criteria

Criteria for inclusion and exclusion are detailed in table1. The PICO framework was as follows:

- Population: Orthodontic patients, patients' clinical images, radiographs, cephalograms.
- Intervention: Orthodontic treatment planning and decision making with AI models. The primary focus will be on interventions having direct clinical significance and effects on the treatment outcome.
- Comparison: Reference standards, conventional treatments, therapeutic consensus.
- Outcomes: Effectiveness and performance of AI models for orthodontic treatment planning and decision-making.

**Table-1: Inclusion and exclusion criteria**

| Inclusion criteria | Exclusion criteria |
|---|---|
| -Articles must be based on artificial intelligence (AI)<br>(AI): will be defined as the self-reported use of AI, deep learning, machine learning, neural network, or any classifier prediction model.<br>-Articles should have a clinical significance in orthodontic interventions.<br>-There should be a mention of some measurable or predictive outcomes that can be quantified.<br>-Articles published from (1st January 2000) until (9th January 2021).<br>-Study design: Meta-analysis, systematic reviews, randomized controlled trial, controlled clinical trials, diagnostic test accuracy studies (DTA: single-gate/two-gate), case-control studies, retrospective and prospective cohorts. | -Articles that focused on areas other than artificial intelligence.<br>-Articles that do not meet the purpose of the review.<br>-Articles with poor insufficient abstract data and whose full text was not available.<br>-Articles in languages other than English and French.<br>-Study design: Narrative reviews, case report, case series, animal studies, in vitro research reports, letters to the editors, commentaries, books, conferences. |

### Information Sources and Literature Search

The search was conducted by two review authors (JM and ID) independently. A combination of controlled vocabulary and medical subject headings (Mesh) terms was elaborated for identifying studies related to this review. The applied restrictions were the publication time, language, and study design (table 1).

Five databases were searched: MEDLINE via PubMed, EBSCO host (Dentistry & Oral Sciences Source database), ScienceDirect, Scopus, and Web of Science (All databases: WOS, KJD, MEDLINE, RSCI, SCIELO).

Furthermore, Open Grey and WorldCat were searched to identify grey literature. A manual search was carried out by scrutinizing references from the included studies, contacting authors, and looking at "related to" or "similar" articles in PubMed. A catch-up search to update the review and identify recently relevant articles was carried out on the 19th of January 2022.

### Study selection

EndNote 20 (Clarivate, Philadelphia, Pa) and Rayyan QCRI [6] were used by 2 review authors (JM and ID) to assist in the study selection process and record decisions, which was conducted in two stages independently: initial screening of titles and abstracts of all studies against the predetermined inclusion criteria, then full text assessment of papers identified as possibly relevant.

Disagreements between the review authors (JM and ID) were discussed, and resolved by consensus after referring to the protocol. However, if deemed necessary, a third person was consulted.

### Data extraction

A customized data collection form was designed for data extraction, which was done independently by two reviewers (JM and ID), and then results were confronted, discussed, and revised together as a team.

### Assessment of the risk of bias in included studies

Two researchers independently assessed the risk of bias of the included articles using "JBI critical appraisal tools" [7]. The potential risk of bias was categorized as low if a study provided detailed information pertaining to 70% or more of the applicable parameters. Moderate risk was considered if a study provided information corresponding to less than 70% to 50% of the applicable parameters, whereas if a study showed missing information regarding more than 50% of the applicable parameters, the study was categorized as exhibiting a high risk of bias.

 

**Certainty assessment**

The "2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence"[8] was used to appraise the level of evidence in included studies. As for the evaluation of quality of evidence and strength of recommendations, "GRADE" (Grading of Recommendations, Assessment, Development, and Evaluation) was used [9, 10].

Decision regarding the quality of evidence and strength of recommendations was carried through a consensus process whereby review authors attributed for each outcome assessed the certainty of evidence using the GRADE methodology.

## STATISTICAL METHODS

Sensitivity (Se), specificity (Sp), likelihood ratio (LR) and diagnostic odds ratio (DOR) are the included metrics for the analysis. First, the random-effects model (DerSimonian and Laird) method for meta-analysis with 95% confidence interval (CI) and a correction factor of 0.5 was employed for descriptive combination of studies and then a hierarchical method: the bivariate model (maximum likelihood) was used for pooling and quantitative combination of studies. Subsequently, a summary receiver operating characteristics (SROC) curve was plotted and publication bias was evaluated using the Deeks' funnel plot asymmetry test.

To test heterogeneity, Chi-square test (Cochrane Q statistic), the Higgins' I2, the $\tau 2$ (Tau2) test, and P values were calculated along with an Interpretation of SROC curve with the 95% prediction region and 95% confidence region. Threshold effect was tested through Spearman's rank correlation coefficient [11].

A p value less than 0.10 on the Q test or an I2 statistic greater than 50% are considered to indicate substantial heterogeneity among DTA study results [12]. Statistical analysis and graphical representations were performed with OpenMeta[Analyst] [13], MetaDTA [14], and Stata 16 (STATA Corp, College Station, Texas, USA).

## RESULTS

**Study selection**

The global search initially yielded 897 records in total. A sum of 234 duplicates were eliminated. Therefore, for the first stage of study selection 663 records were screened by titles and abstracts, discarding in the process 646 articles. After meticulous reading and discussion, 10 studies were included and the remaining articles were excluded.

A catch-up search was executed in the PubMed database on the 19$^{th}$ of January 2022 (from the 9$^{th}$ of January 2021 till the 19$^{th}$ of January 2022), using the same search query. This search yielded 140 records. Only two articles made it to final inclusion, resulting in a total of 12 articles.

The global selection process is illustrated in the PRISMA 2020 flow diagram (Figure 1).

**Study characteristics**

Ten studies were diagnostic test accuracy studies (DTA); from which nine had single-gate case control design [15-23] and one study [24] had a single-gate cross sectional design. One study [25] was a recent systematic review (SR) and one study [26] a retrospective cohort (table 2).

More than 66% of studies identified were published in the last four years and most were conducted in South Korea (5 out of 12). The sample size in all trials ranged from 56[20] to 1000 [24] data sets with a total of 4370 patient records (table 3).

Age range across individual studies was from 6.3 to 52 years of age, and a mean age of 19.48 years. Diversified AI approaches were used: Fuzzy modelling, artificial neural networks (ANNs), bayesian networks (BNs), convolutional neural networks (CNNs) and machine learning (ML): Boruta method, XGBoost classifier, neural network model, random forest classifier. ANNs was the most used model [16,17,19,23], followed by BNs [24,26] and CNNs [18, 21].

The focus of all interventions was orthodontic treatment planning and decision-making with slightly varying study factors. All comparisons were reference standards executed by experienced clinicians apart from two studies (Xie 2010: DTA) (Nieri 2010: cohort) that did not mention any comparison details. The number of experienced specialists ranged from 1 to 8 that had on average 12.39 years of experience.

The evaluation of AI models' performance was reported through: examiners' agreement and average satisfaction, success rates of the AI model, ICC value, accuracy, sensitivity, specificity, AUC, F1-score, and 10-fold cross validation accuracy.

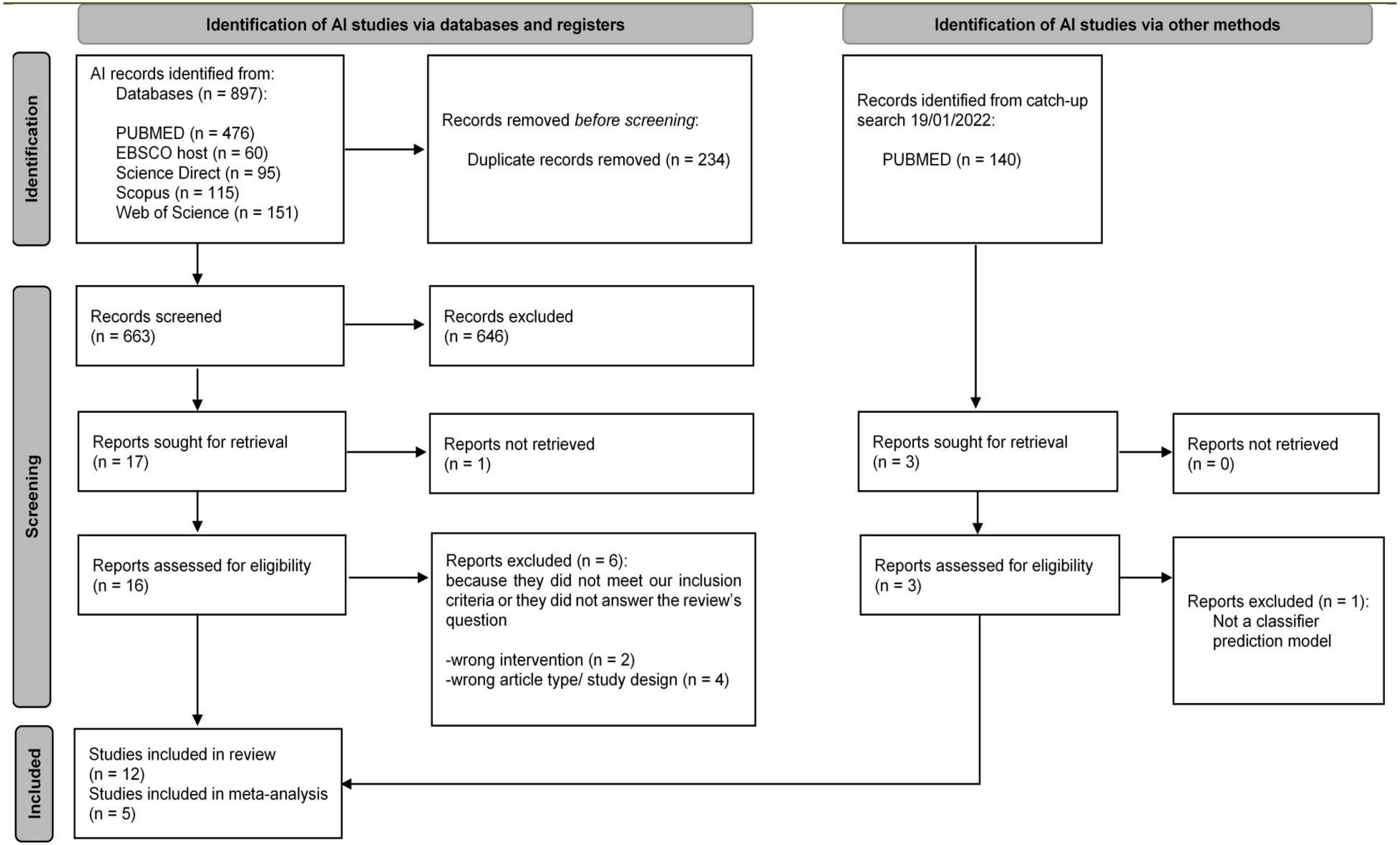


**Fig-1: PRISMA 2020 flow diagram.**

**Table-2: Articles' identification**

| Study ID and Title | First author | Year | country | Study design | Journal | Aim of study |
|---|---|---|---|---|---|---|
| Akçam 2002 [15]<br>Fuzzy modelling for selecting headgear types | M. Okan Akçam | 2002 | Japan | DTA single gate case control | European Journal of orthodontics | To develop a computer-assisted inference model for selecting appropriate types of headgear appliance for orthodontic patients and to investigate its clinical versatility as a decision-making aid for inexperienced clinicians. |
| Choi 2019 [16]<br>Artificial intelligent model with neural network machine learning for the diagnosis of orthognathic surgery | Hyuk-Il Choi | 2019 | Korea | DTA single gate case control | The journal of craniofacial surgery | To develop a new artificial intelligent model for surgery/non-surgery decision and extraction determination, and to evaluate the performance of this model. |
| Jung 2016 [17]<br>New approach for the diagnosis of extractions with neural network machine learning | Seok-Ki Jung | 2016 | Korea | DTA single gate case control | American journal of orthodontics and dentofacial orthopedics | To construct an artificial intelligence expert system for the diagnosis of extractions using neural network machine learning and to evaluate the performance of this model. |
| Khanagar 2020 [25]<br>Scope and performance of artificial intelligence technology in orthodontic diagnosis, treatment planning, and clinical decision-making – A systematic review | Sanjeev B. Khanagar | 2020 | Saudi Arabia | Systematic review | Journal of dental sciences | To document the scope and performance of the artificial intelligence-based models that have been widely used in orthodontic diagnosis, treatment planning and predicting the prognosis. |
| Kim 2021 [18]<br>Influence of the depth of the convolutional neural networks on an artificial intelligence model for diagnosis of orthognathic surgery. | Ye-Hyun Kim | 2021 | Korea | DTA single gate case control | MDPI Journal of personalized medicine | To investigate the relationship between image patterns in cephalometric radiographs and the diagnosis of orthognathic surgery and propose a method to improve the accuracy of predictive models according to the depth of the neural networks. |
| Li 2019 [19]<br>Orthodontic treatment planning based on artificial neural networks | Peilin LI | 2019 | China | DTA single gate case control | Scientific reports | To use a multilayer perceptron artificial neural networks to predict orthodontic treatment plans, including the determination of extraction-nonextraction, extraction patterns, and anchorage patterns. |
| Lin 2020 [20]<br>Early prediction of the need for orthognathic surgery in patients with repaired unilateral cleft lip and palate | Guang Lin | 2020 | Korea | DTA single gate case control | Thesis Reference published ahead of print 2020 | To determine the cephalometric parameters that can predict the future need for orthognathic surgery or distraction osteogenesis (DO) in Korean patients with |

| | | | | | | |
|---|---|---|---|---|---|---|
| using machine learning and longitudinal lateral cephalometric analysis data | | | | | Article published in 2021 (The journal of craniofacial surgery) | repaired unilateral cleft lip and palate (UCLP) by using machine learning and longitudinal lateral cephalometric analysis. |
| Nieri 2010 [26] Factors affecting the clinical approach to impacted maxillary canines: A Bayesian network analysis | Michele Nieri | 2010 | USA | Retrospective cohort | American journal of orthodontics and dentofacial orthopedics | To apply Bayesian networks to evaluate the relative role and possible causal relationships among various factors affecting the diagnosis and final treatment outcome of impacted maxillary canines. |
| Shin 2021 [21] Deep learning based prediction of necessity for orthognathic surgery of skeletal malocclusion using cephalogram in Korean individuals. | WooSang Shin | 2021 | Korea | DTA single gate case control | BMC oral health | To develop a deep learning network to automatically predict the need for orthognathic surgery using cephalogram. |
| Suhail 2020 [22] Machine learning for the diagnosis of orthodontic extractions: A computational analysis using ensemble learning | Yasir Suhail | 2020 | USA | DTA single gate case control | MDPI Bioengineering | To create an artificial intelligence decision-making model for the diagnosis of extractions using neural network machine learning. |
| Thanathornwong 2018 [24] Bayesian-based decision support system for assessing the needs for orthodontic treatment | Bhornsawan Thanathornwong | 2018 | Thailand | DTA single gate cross sectional | Healthcare informatics research | To develop a clinical decision support system to help general practioners access the need for orthodontic treatment in patients with permanent dentition. |
| Xie 2010 [23] Artificial neural network modelling for deciding if extractions are necessary prior to orthodontic treatment | Xiaoqiu Xie | 2010 | China | DTA single gate case control | Angle Orthodontist | To construct a decision-making expert system (ES) for the orthodontic treatment of patients between 11 and 15 years old to determine whether extraction is needed by using artificial neural networks (ANN). Specifically, uncovering the factors that affect this decision-making process. |

**Table-3: Epidemiological and clinical data**

| Study ID | Population (Sample) | Intervention | | | comparison | Evaluation | Results/Outcomes (+) effective, (-) non effective, (N) neutral | Conclusions |
|---|---|---|---|---|---|---|---|---|
| | | Study factor | Modality | AI approach | | | | |
| Akçam 2002[15] | **85** orthodontic patients' pre-treatment records<br>Mean age:<br>12.9±4.6 years<br>Age range:<br>[8.1;31.1] years<br>Males: 33<br>Females: 52 | Headgear type:<br>-Low<br>-Medium<br>-High pull | -Dental casts<br>-Intra and extra-oral photographs<br>-Panoramic radiographs<br>-Lateral cephalograms | Fuzzy modelling | 8 experienced orthodontists (6 men, 2 women)<br>Experience:<br>-Mean:<br>14.7±3.7 years<br>-Range:<br>[10.1;20.9] years | Average satisfaction was 95.6% (SD 2.6) | (+) effective<br>-All of the examiners exceeded a kappa score of 0.7, allowing them to evaluate the inference model<br>-The average satisfaction rate of the examiners was 95.6%, and for 83 out of the 85 cases, 97.6%. | -The majority of the examiners were satisfied with the recommendations of the system.<br>-The system developed was reliable and effective for clinical use in orthodontics. |
| Choi 2019[16] | **316** cases:<br>-160 surgical cases<br>-156 non-surgical<br>*204 learning set:<br>-136 training set<br>-68 validation set<br>*112 test set<br>Mean age (M)<br>22.1±4.8 years<br>Mean age (F)<br>23.6±6.5 years<br>Males: 123<br>Females: 193 | -Tooth malocclusion<br>-Orthognathic surgery planning | -Lateral cephalograms | ANNs | 1 orthodontist<br>Experience:<br>10 years | -Success rate of surgery/non-surgery diagnosis: 96%<br>-Success rate of detailed diagnosis: 91%<br>-ICC value:<br>[0.97;0.99] | (+) effective<br>-The success rate of surgery /non-surgery diagnosis was 96% for the total.<br>-Class II and III surgical type classification was 100% successful in all sets.<br>-For surgical cases, the success rate of extraction diagnosis for Class II surgery was 97%. While for Class III surgery it was 88%.<br>-the final diagnosis success rate was 91%. | The neural network machine learning artificial intelligent model was useful for the diagnosis of orthognathic surgery cases. |
| Jung 2016[17] | **156** cases<br>*96 learning set:<br>-64 training set<br>-32 validation set<br>*60 test set<br>Mean age (M)<br>23 years<br>Mean age (F)<br>25 years<br>Males: 62<br>Females: 94 | -Tooth malocclusion<br>-Extraction planning | -Lateral cephalograms | ANNs | 1 orthodontist<br>Experience:<br>10 years | -Success rates of extraction/ non-extraction diagnosis: 93%<br>-Success rate of detailed diagnosis of the extraction patterns: 84%<br>-ICC value:<br>[0.97;0.99] | (+) effective<br>-The decision-making success rate was 93% for the diagnosis of extraction vs nonextraction.<br>-In the diagnosis of identical vs differential extraction, the success rate was 89%.<br>-The final success rates were 85% in the learning set, 82% in the test set, and 84% in total. | -ANN based AI expert systems could be useful in orthodontics.<br>-Proper selection of the input data, appropriate organization of the modeling, and preferable generalization Improved the performance of the model. |
| Kim 2021[18] | **960** cases<br>*640 no surgery<br>*320 surgery<br>-training set: 810<br>-test set: 150<br>Males: 468 | -Tooth malocclusion<br>-Orthognathic surgery planning | -Cephalograms | CNNs: ResNet-18, 34, 50, and 101 | 1 orthodontist<br>Experience:<br>Not mentioned | Best performance was achieved by ResNet-18:<br>-**AUC:** 0.979<br>-**Accuracy:** 0.938<br>-**Sensitivity:** 0.882 | (+) effective<br>-Accuracy in the test set for the ResNet-18, 34, 50, and 101 was 93.80%, 93.60%, 91.13%, and 91.33%, respectively.<br>-In screening, ResNet-18 had the | -The developed models were successful in diagnosing the need for orthognathic surgery.<br>-The ResNet-18 attained the highest |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Females: 492<br>Mean age:<br>24.6 years SD: 4.9<br>Age range:<br>[15;37] years | | | | | -**Specificity**: 0.966 | best performance with an AUC of 0.979, followed by ResNets-34, 50, and 101 at 0.974, 0.945, and 0.944, respectively. | performance among the other developed models with a success rate of 93.80%. |
| Khanagar 2020[25] | **16** research articles<br>[2009-2019] | -Diagnosis<br>-Treatment planning<br>-Predicting prognosis | -Patients clinical images<br>-Radiographs<br>-Cephalograms involving oral and maxillofacial structures | AI based models | -Expert opinions<br>-Reference standards | Measurable or predictive outcomes such as:<br>-Accuracy<br>-Sensitivity<br>-Specificity<br>-ROC<br>-AUC<br>-ICC | (+) effective<br>AI technology was extensively applied for determining need for orthodontic treatment needs and extractions, identifying cephalometric landmarks, determining the degree of maturation of the cervical vertebra, and predicting the facial attractiveness. Most used AI models were ANNs or CNNs. | -AI models have performed exceptionally well, with an accuracy and precision similar to the trained examiners.<br>-These systems can be of great value in orthodontics. |
| Li 2019[19] | **302** cases:<br>-222 extraction<br>-80 non-extraction<br>-182 training set<br>-60 validation set<br>-60 test set<br>Mean age:<br>17.16±5.71 years<br>Age range:<br>[9;40] years | -Tooth malocclusion<br>-Extraction planning | Medical records before orthodontic treatment were collected:<br>-demographic information<br>-extraoral photos<br>-intraoral photos<br>-pre-treatment dental casts<br>-lateral cephalometric measurements | ANNs | 2 orthodontists<br>Experience:<br>-26 years<br>-12 years | -**Accuracy:**<br>-Extraction/no extraction: 94%<br>-Extraction patterns: 84.2%<br>-Anchorage patterns: 92.8%<br>-**Sensitivity:**<br>94.6%<br>-**Specificity:**<br>93.8%<br>-**AUC:**<br>0.982 (95% CI 0.968–0.995) | (+) effective<br>-to enhance applicability, the model suggests several practicable alternatives for doctors to choose from to compensate for the decision-making variability on extraction patterns.<br>-The most important features for prediction of the ANN are "crowding, upper arch" "ANB" and "curve of Spee". | The developed ANN model was useful for providing good guidance for orthodontic treatment planning for less-experienced orthodontists. |
| Lin 2020[20] | **56** cases:<br>-10 surgical<br>-46 non-surgical<br>Males: 31<br>Females: 25<br>Mean age:<br>(T0): 6.3 years<br>(T1): 16.7 years | -Orthognathic surgery planning<br>-Unilateral cleft lip and palate (UCLP) | -Lateral cephalograms<br>At T0 and T1<br>(T0): before orthodontic / orthopedic treatment<br>(T1): at least 15 years of age | Machine learning (ML):<br>-Boruta method<br>-XGBoost classifier | -1 orthodontist<br>-1 surgeon | -**10-fold cross validation Accuracy:**<br>87.4%<br>-**Sensitivity:**<br>97.83%<br>-**Specificity:**<br>90.00%<br>-**F1-score:**<br>0.714<br>-**A 2x2 confusion matrix** | (+) effective<br>The following indices: ANB, PP-FH, CF, and facial convexity angle were determined as predictive parameters of the future need for orthognathic surgery.<br>The developed model had a 10-fold cross-validation accuracy of 87.4% with an F1-score of 0.714. | At age of 6 years, determining the future need for orthognathic surgery in patients with UCLP using cephalometric predictors was possible with a good accuracy. |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Nieri 2010[26] | **168** patients :<br>-125 unilateral impaction<br>-43 bilateral impaction<br>Males: 40<br>Females: 128<br>Mean age:<br>17.2±6 years<br>Age range:<br>[12.8;52.0] years<br>Follow-up: 17 years | -Tooth malocclusion<br>-Impacted maxillary canines | Demographic, orthodontic, and periodontal variables | BNs | Not applicable | Not applicable | (+) effective<br>-168 impacted canines were successfully moved and aligned.<br>-The BN analysis determined that bilateral impaction was associated with palatal impaction and longer treatment time.<br>-The pre-treatment α-angle was an important factor for the duration of orthodontic traction. | -Bayesian network analysis was useful to identify possible relationships among the variables considered for diagnosis and treatment of impacted canines. |
| Shin 2021[21] | **840** cases<br>*622 no surgery<br>*218 surgery<br>-Training set: 273 surgery<br>98 no surgery<br>-Validation set:<br>30 surgery<br>11 no surgery<br>-Test set:<br>304 surgery<br>109 no surgery<br>Males: 461<br>Females: 379<br>Mean age:<br>23.2 years<br>Age range:<br>[19;29] years<br>SD: 3.15 | -Tooth malocclusion<br>-Orthognathic surgery planning | -Transverse and longitudinal cephalograms | CNNs | 6 specialists:<br>-2 orthodontists<br>-3 maxillofacial surgeons<br>-1 maxillofacial radiologist<br><br>Experience:<br>Not mentioned | -**Accuracy:**<br>0.954<br>-**Sensitivity:**<br>0.844<br>-**Specificity:**<br>0.993<br>-**A 2x2 contingency table** | (+) effective<br>-In the test set, 394 out of a total of 413 were properly classified.<br>-The accuracy of the developed model reached 95.4%. | CNN was useful for determining the need for orthognathic surgery. |
| Suhail 2020[22] | **287** pre-treatment patient records | -Tooth malocclusion<br>-Extraction planning | Medical charts and conventional diagnostic records:<br>-lateral head films (cephalometric X-rays)<br>-panoramic radiographs<br>-facial photographs<br>-intraoral photographs | Machine learning (ML):<br>-neural network model<br>-random forest ensemble classifier | 5 orthodontists<br><br>Experience:<br>Average: 9 years | The out-of-bag accuracy:<br>ranged between 60% and 75%. | (+) effective<br>-The agreement between the experts on the primary outcome of treatment varied from 65% to 71%.<br>-Agreement on either the primary or alternative outcome varied from 93% to 98%.<br>-The random forest classifier performed better than the neural network model for the prediction of the specific extraction treatment. | A random forest ensemble classifier was useful for extraction planning with high performance, within the range of the inter-expert agreement. |
| Thanathornwong 2018[24] | **1) 1000 data-sets**<br>Mean age:<br>23.2 years<br>Age range:<br>[19;29] years<br>SD: 3.15<br>Males: 375 | -Tooth malocclusion<br>-Treatment needs | -Upper and lower arch impressions<br>-Photographs | BNs | 2 orthodontists<br><br>Experience:<br>At least 5 years<br><br>Assessed orthodontic | -**Accuracy:**<br>0.96<br>-**Sensitivity:**<br>0.95<br>-**Specificity:**<br>1.00<br>-**AUC:** | (+) effective<br>-The two orthodontists had a high degree of agreement in their diagnoses and their judgements regarding the need for orthodontic treatment.<br>-The decision support system had a | -The BNs system exhibited high performance and promising results.<br>-The model had a high degree of accuracy in classifying patients into |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Females: 625<br>Mean age:<br>17.4±2.51 years<br>**2) 20 new patients**<br>-evaluation set<br>Males: 5<br>Females: 15<br>Age range:<br>[14;19] years | | | | treatment needs for the 20 new patients | 0.91 | high degree of agreement with the two orthodontists. | groups needing and not needing orthodontic treatment. |
| Xie 2010[23] | 200 cases:<br>-120 extraction<br>-80 non-extraction<br>-180 training set<br>-20 testing set<br>Age range:<br>[11;15] years | -Tooth malocclusion<br>-Extraction planning | -Cast measurements<br>-Lateral cephalograms | ANNs | Not mentioned | -**Accuracy:**<br>80% | (+) effective<br>-In the test set, the ANN accuracy was 80%.<br>-For determining extraction vs non-extraction, ''Anterior teeth uncovered by incompetent lips'' and ''IMPA (L1-MP)'' were the two indices that gave the biggest contributions to the ANN, While, the index of FMA (FH-MP) gave the smallest contribution. | -The developed ANN was effective in determining the need for extractions for malocclusion patients between 11 and 15 years old, with high accuracy.<br>-When the practitioner is deciding orthodontic extractions, the indices: ''anterior teeth uncovered by incompetent lips'' and ''IMPA (L1-MP)'' must be considered. |

## Risk of bias in included studies

Two DTA studies [15, 23] had high risk of bias, five DTA studies [16–19, 22] had moderate risk of bias, and three DTA studies [20, 21, 24] had low risk of bias, while the SR [25] and the cohort [26] Studies had both low risk of bias (table 4).

**Table-4: Risk of bias in included studies**

| Checklist / Study ID | Checklist for diagnostic test accuracy studies | | | | | | | | | | | Assessment | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Q1** | **Q2** | **Q3** | **Q4** | **Q5** | **Q6** | **Q7** | **Q8** | **Q9** | **Q10** | | | |
| Akçam 2002[15] | U | N | U | N | U | Y | Y | U | Y | Y | | **40%** | - |
| Choi 2019[16] | U | N | Y | U | Y | Y | Y | U | Y | Y | | **60%** | = |
| Jung 2016[17] | U | N | Y | U | Y | Y | Y | U | Y | Y | | **60%** | = |
| Kim 2021[18] | U | N | U | N | Y | Y | Y | Y | Y | Y | | **60%** | = |
| Li 2019[19] | U | N | Y | U | Y | Y | Y | U | Y | Y | | **60%** | = |
| Lin 2020[20] | U | N | Y | N | Y | Y | Y | Y | Y | Y | | **70%** | + |
| Shin 2021[21] | Y | N | Y | N | Y | Y | Y | Y | Y | Y | | **80%** | + |
| Suhail 2020[22] | U | N | Y | N | U | Y | Y | U | Y | Y | | **50%** | = |
| Thanathornwong 2018[24] | U | Y | Y | Y | Y | Y | Y | Y | Y | Y | | **90%** | + |
| Xie 2010[23] | U | N | Y | N | Y | U | Y | NA | U | Y | | **44.4%** | - |
| | **Checklist for cohort studies** | | | | | | | | | | | | |
| | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Q11 | | |
| Nieri 2010[26] | NA | NA | Y | N | N | Y | Y | Y | Y | NA | Y | **75%** | + |
| | **Checklist for systematic reviews** | | | | | | | | | | | | |
| | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Q11 | | |
| Khanagar 2020[25] | Y | Y | Y | Y | Y | Y | U | Y | NA | Y | U | **80%** | + |

Y: yes, N: no, U: unclear, NA: not applicable, +: low risk, -: high risk, =: moderate risk;

**Certainty assessment**

Ten articles had an OCEBM level of 4 and a GRADE recommendation of 1-C, this is because these articles were DTA case-controls in nature and it was difficult to draw recommendations for practice from them. Only one DTA study [24] had an OCEBM level of 2 because it had a cross-sectional design. On the other hand, the SR [25] included had a high quality with an OCEBM level of 1 and a GRADE recommendation of 1-B (table 5).

**Table-5: Certainty assessment**

| Study ID | OCEBM | GRADE |
|---|---|---|
| Akçam 2002 [15] | 4 | 1-C |
| Choi 2019 [16] | 4 | 1-C |
| Jung 2016 [17] | 4 | 1-C |
| Khanagar 2020 [25] | 1 | 1-B |
| Kim 2021 [18] | 4 | 1-C |
| Li 2019 [19] | 4 | 1-C |
| Lin 2020 [20] | 4 | 1-C |
| Nieri 2010 [26] | 4 | 1-C |
| Shin 2021 [21] | 4 | 1-C |
| Suhail 2020 [22] | 4 | 1-C |
| Thanathornwong 2018 [24] | 2 | 1-C |
| Xie 2010 [23] | 4 | 1-C |

**META-ANALYSIS FINDINGS**

Only five studies [18–21,24] were included in the meta-analysis, since in most studies the raw data necessary to meta-analyse diagnostic accuracy measures were unavailable.

Se, Sp, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and DOR forest plots were generated (figure2).

All studies exhibited a PLR>5 but with varying values, as the highest estimate PLR value was 381.67[24]. While for the other studies the values ranged from 6.37[21] to 29.33[18].

As for NLR, all studies had a value less than 0.2. The lowest NLR was 0.008[21], whereas the highest value was 0.124[18] (figure 2).

Pooled sensitivity, specificity, False positive rate (FPR), DOR, and area under the curve (AUC) with 95% CI of AI models' performance were: 0.965 (95% CI 0.921-0.985), 0.962 (95% CI 0878-0.989), 0.038 (95% CI 0.011-0.122), 695.537 (95% CI 232.742-2078.572), 0.99 (95% CI 0.98 - 1.00), respectively.

The higher amount of DOR is indicative of the fact that the approaches can determine the right treatment plan with high overall accuracy.

Pooled analysis of the crude value of TP, FP, FN, and TN revealed that the accuracy of the AI algorithms reached 95.47%.

Pooled PLR was 25.304 (95% CI 7.686-83.310)>5 and the NLR was 0.036 (95% CI 0.016-0.081)<0.2, which indicates a clinically useful test and strong diagnostic evidence.

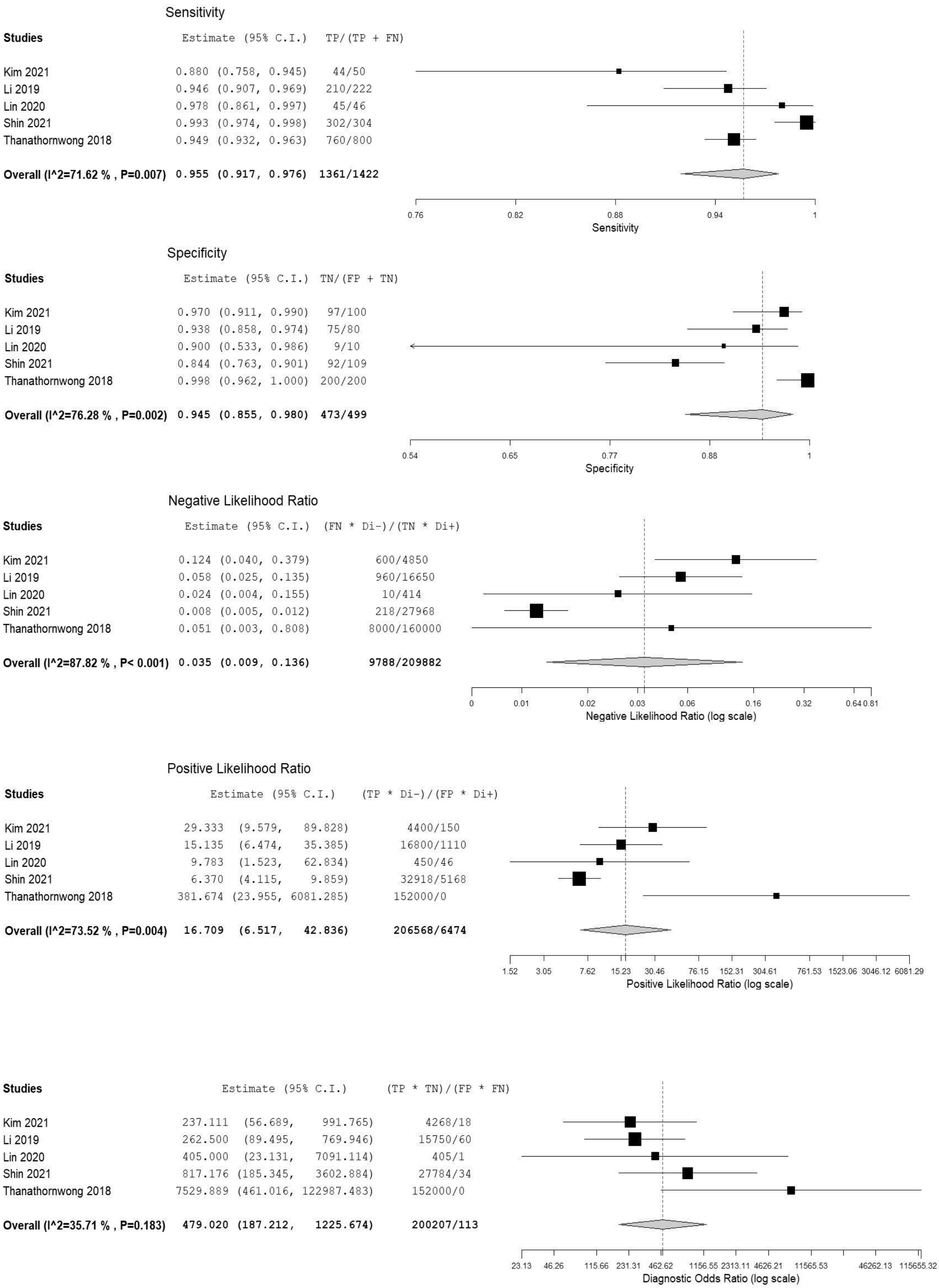


**Fig-2: Se, Sp, LRs and DOR forest plots**

**Assessment of heterogeneity**

There was a noticeable heterogeneity in NLR, PLR, Sp and Se (I2: 87.82, 73.52, 71.62, and 76.28 respectively). While the odds ratio showed low heterogeneity (I2: 35.70, p>0.1).

Heterogeneity is indicated by how closely the included data fits to an SROC curve. Data that fits a typical shoulder-shaped SROC curve tightly indicates low heterogeneity [11]. The ROC plot, shows that this criterion is not satisfied, as one study [24] seems to be astray from the curve made by other studies (Figure 3).

In the SROC space, the 95% prediction region is much larger than the 95% confidence region and the SROC curve does not seem to include all studies which is an indicator of heterogeneity (Figure 3).

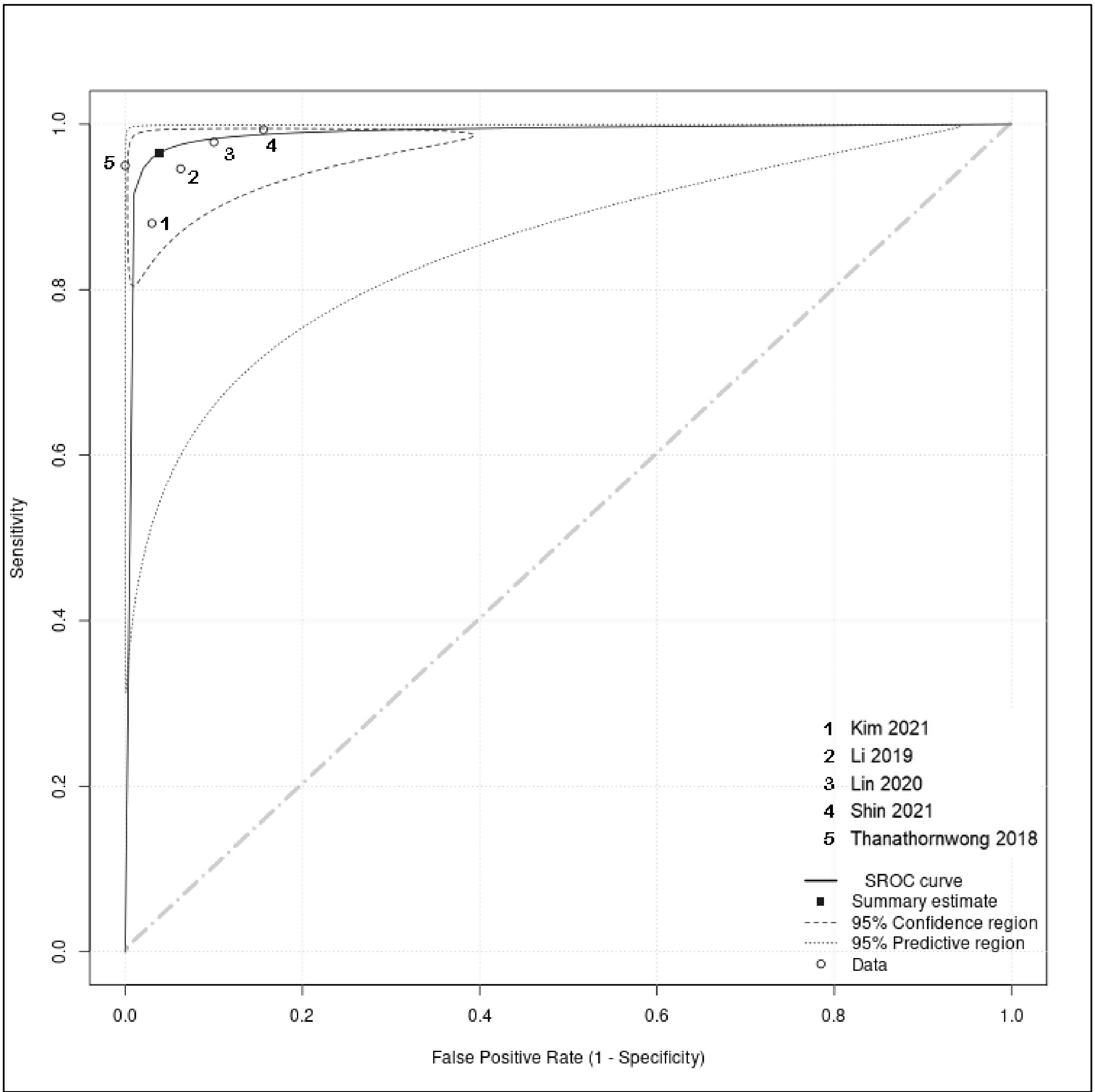


**Fig-3: Summary receiver operating characteristic curve with 95% confidence region and prediction region.**

A sensitivity analysis on the SROC curve was conducted to further assess the implication of (Thanathornwong 2018) in overall heterogeneity. As can be seen in the bivariate model SROC curve (Figure 4) after excluding this study [24], the 95% prediction region and the 95% confidence region are tightly fit compared to the original model and the included data from other studies fit closely in the SROC curve. Thus, (Thanathornwong 2018) is a source of heterogeneity.

The sensitivity-specificity dependency based on threshold variability can be assessed using

Spearman's rank correlation coefficient. The threshold effect is regarded as substantial if a significant correlation exists, with a correlation coefficient of 0.67 or higher [11]. The determined value (0.44) was comprised between 0.36 - 0.67 and thus, indicating a moderate threshold effect.

Therefore, the presence of two different heterogeneity causes; the first one being the threshold effect and the second one being the heterogeneity caused by the study of Thanathornwong [24].

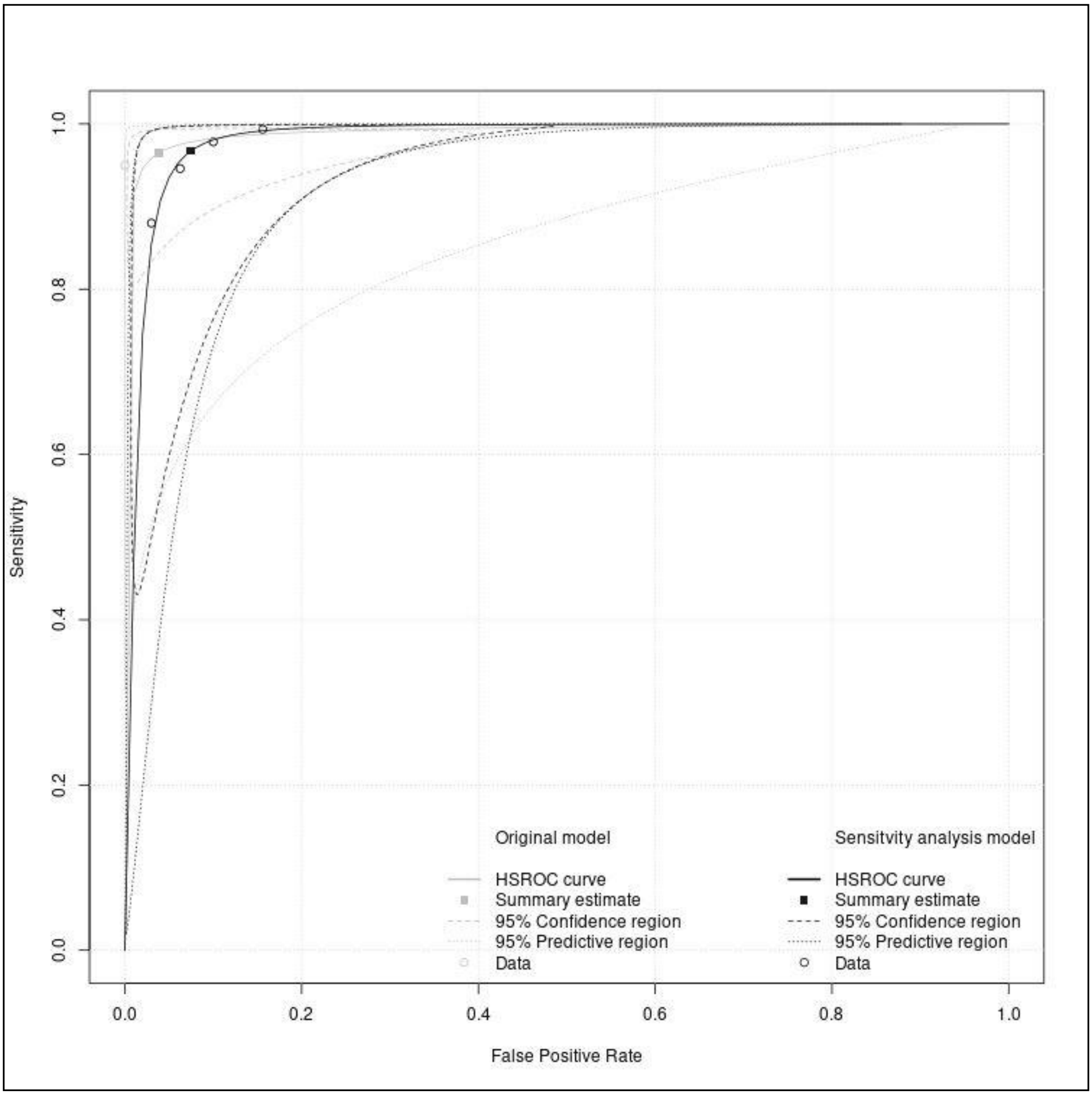


**Fig-4: Summary receiver operating characteristic curve sensitivity analysis**

**Assessment of publication bias**

The Deeks' funnel plot of studies exhibited a grossly symmetrical shape with respect to the regression line (Figure 5), and the asymmetry test showed no apparent evidence of publication bias (p=0.20).

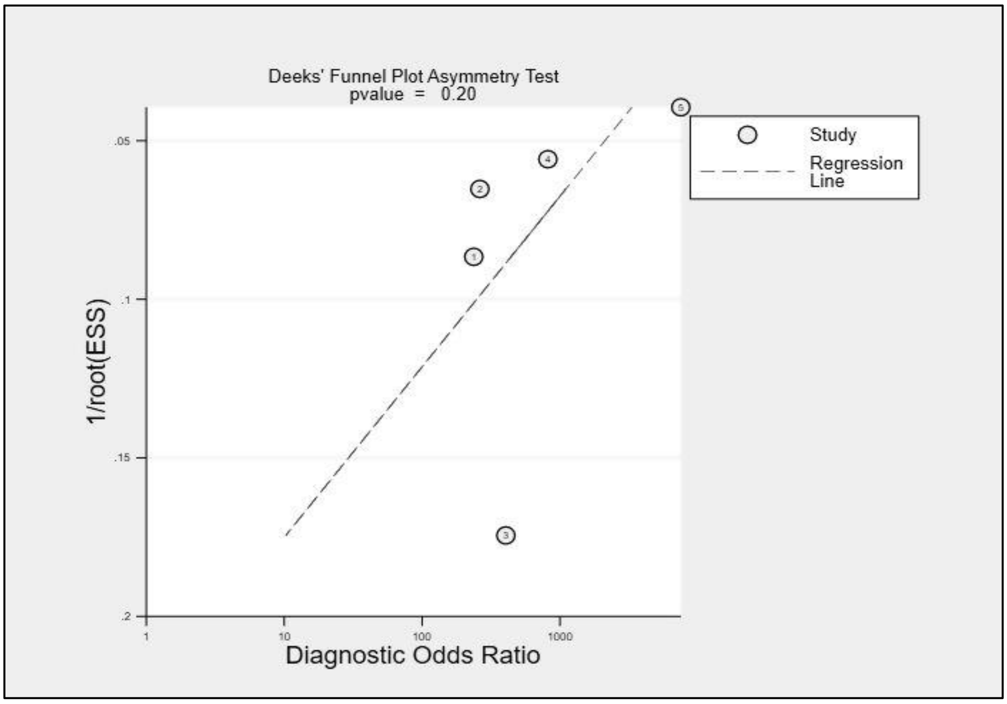


**Fig-5: Deek's funnel plot of studies**

**Summary of findings**

The GRADE summary of findings table was done through GRADEpro GDT (GRADEpro GDT: GRADEpro Guideline Development Tool, 2020). For assessing the certainty of the body of evidence across outcomes, the study design was set as case-control type accuracy study, the risk of bias and inconsistency as serious, indirectness and imprecision as not serious, publication bias as undetected and effect was determined as large (table 6).

The values of pooled sensitivity and specificity of the five studies indicated that the overall rates of correct predictions of orthodontic treatment plans were high. Either in predicting orthodontic treatment, extractions or orthognathic surgery needs, the AI models exhibited high predictive values and good discriminative power for patients' classification with small error margins.

**Table-6: Grade summary of findings**

**What is the effectiveness and performance of AI models for orthodontic treatment planning and decision-making?**
**Patient or population**: Orthodontic patients' records
**Settings:** Orthodontic clinic / departments
**New test**: AI models treatment planning and decision-making systems
**Reference test**: Reference standards (by experienced clinicians)
**Pooled sensitivity**: 0.96 (95% CI: 0.92 to 0.98) | **Pooled specificity**: 0.96 (95% CI: 0.88 to 0.99)

| **Test result** | **Number of results per 1,000 patients tested (95% CI)** | | | **Number of participants (studies)** | **Certainty of the Evidence (GRADE)** |
|---|---|---|---|---|---|
| | **Prevalence 20% typically in orthognathic surgery needs** | **Prevalence 50% typically in orthodontic treatment needs** | **Prevalence 40% typically in extraction needs** | | |
| **True positives** | **193** (184 to 197) | **483** (461 to 493) | **386** (368 to 394) | 1921 (5) | ⊕⊕⊕◯ **Moderate** |
| **False negatives** | **7** (3 to 16) | **17** (7 to 39) | **14** (6 to 32) | | |
| **True negatives** | **770** (702 to 791) | **481** (439 to 495) | **577** (527 to 593) | 1921 (5) | ⊕⊕⊕◯ **Moderate** |
| **False positives** | **30** (9 to 98) | **19** (5 to 61) | **23** (7 to 73) | | |

**CI:** confidence interval

# DISCUSSION

The field of orthodontics primarily deals with the diagnosis of malocclusion and planification of an organized, customized treatment [27].

Treatment planning in orthodontics should maximize the benefits to the patient and minimize the associated risks. To better ensure a suitable treatment plan, there should be a rational decision-making process made through diagnostic tests and based on each patient's case. Even so, the 'perfect' treatment plan specific to each patient is only relative, as reference standards in orthodontics are mainly executed by humans, and it is based on the experts' clinical experience. As a matter of facts, orthodontists' treatment plan can vary for a specific case [28]. The AI systems deal with computational based automated models that can think and act rationally, thus, decreasing the likelihood of human subjectivity during the decision-making process. The AI-based models assist healthcare professionals in enhancing patient care. They can help clinicians operate more efficiently by saving time and suggesting therapeutic options that the practitioner had not considered.

Throughout the decision-making process, orthodontists are usually confronted with many variables and need to rely on heuristics to produce efficient decisions based on confounding and limited information. In practical situations distinguished by excessive aspects of variability and uncertainty, cognitive biases and judgment errors related to heuristics are common. The current review assessed three studies [15, 24, 26] in this area that used fuzzy modelling and BNs.

BNs seems to be the most suitable AI approach to deal with uncertainty and determine causal relationships between variables even in the case of missing clinical data. The results imply that these models may be used as a tool for less experienced orthodontists to predict the need for orthodontic treatment and treatment planning, as well as a useful tool for secondary opinion.

Covering more than two decades of research, we found that recently AI expert systems have been used on deciding the need for orthodontic extractions and the extraction patterns. Four included studies covered this section. ANNs seems to be the most used AI model in orthodontic extractions planning as three studies used ANNs [17, 19, 23] and one study [22] used ML neural network model and random forest ensemble classifier. Only one study [23] evaluated the need for extractions alone, while other studies reported the detailed extraction diagnosis. All AI systems were judged as effective.

The results obtained from these studies suggest that the AI expert systems can be useful for clinical decision making. These pilot studies' results are promising and suggest that there is more room for improving these models.

Orthognathic surgery can drastically change appearance and occlusal function and thus, impacting the patient's sense of self and well-being. Like orthodontic extractions, surgery is irreversible, and its huge impacts should be assessed with care before carrying it out on the patient. In this context, this review included four studies[16,18,20,21] that used ANNs, CNNs and ML algorithms. These models performed well in orthognathic surgery planning, with high accuracies.

Delivering a customized precise treatment for each patient has always been one of the important challenges facing practitioners. AI technology drives us closer to overcome this hurdle. With the tremendous amount of diversified clinical data stored in its databases, AI-based systems can be used as advisory tools for less experienced orthodontists and those in training. Thus, procuring a secondary opinion that can help practioners achieve successful orthodontic treatments, detailed diagnoses, and accurate treatment plans with adequate outcomes. Which will ultimately save time and resources and help responding to the needs of society.

**Quantitative synthesis analysis**

All the studies [18–21] included in the meta-analysis were DTA single gate case-controls, apart from one study [24] that had a cross sectional design, but only the original internal dataset used for making the AI system was included in the meta-analysis. Thus, all studies were considered case-controls.

Three studies [20, 21, 24] had low risk of bias. While the other two [18,19] had moderate risk of bias. Since CNNs is just one kind of ANNs, then three studies [18, 19, 21] were using ANNs and two studies [20, 24] using BNs and ML algorithms, which can be considered as a source of heterogeneity.

(Thanathornwong 2018) was a source of heterogeneity due to its design, sample size, different AI algorithm, or the application of the correction factor 0.5 due to its specificity=1.

There are noticeable sources of heterogeneity between studies such as the study factor, modality, and AI approach. Also, the reference standard comparison was not consistent across studies.

The goal was to have a global view on the performance of the developed AI models in the orthodontic field, especially in treatment planning and decision-making, which was satisfied.

The findings indicate high AI models' performance despite the clear limitations in the studies included. Thus, it can only be concluded that AI models were successful in predicting valid treatment plans with accurate decisions. These models can be further improved for more applicable consistent results.

## RECOMMENDATIONS

The future use of available high accuracy AI expert systems as a diagnostic aid and a clinical decision support system is advised, since practioners can refer to them for a second opinion.

To assess clinical effectiveness and the practical utility of AI models as diagnostic tests compared to the conventional reference standards, diagnostic test randomised controlled trials (D-RCTs) should be carried.

Researchers are encouraged to publish the code of developed AI systems as open source, so other researchers can work on improving the existing models and collaborate to enhance the systems' accuracy and applicability.

## CONCLUSION

The impact of artificial intelligence is undeniable as AI technology is able to ameliorate the diagnostic reliability and precision for orthodontic treatments, therewith assisting the clinicians in operating more accurately and efficiently.

## ACKNOWLEDGEMENTS

All authors declare that there is no conflict of interest related to this work. This research did not receive any financial funds. The following article was adapted from the doctoral thesis of Dr. Jihed M'hamed.

## REFERENCES

1. Alotaibi, Y. K., & Federico, F. (2017). The impact of health information technology on patient safety. *Saudi medical journal*, *38*(12), 1173.
2. Joiner, I.A. (2021). Artificial Intelligence. Emerg Libr Technol [Internet]. 2018 [cited 2022 Mar 12];1–22. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780081022535000022
3. Davenport, T., Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Futur Healthc J [Internet]. 2019 Jun [cited 2022 Mar 12];6(2):94. Available from: /pmc/articles/PMC6616181/
4. Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D. (2022). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ [Internet]. 2021 Mar 29 [cited 2022 Mar 10];372. Available from: https://www.bmj.com/content/372/bmj.n71
5. Cumpston, M., Li, T., Page, M. J., Chandler, J., Welch, V. A., Higgins, J. P., & Thomas, J. (2019). Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. *Cochrane Database Syst Rev*, *10*(ED000142).
6. Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic reviews*, *5*(1), 1-10.
7. Critical-appraisal-tools - Critical Appraisal Tools | Joanna Briggs Institute [Internet]. [cited 2022 Mar 10]. Available from: https://jbi.global/critical-appraisal-tools
8. OCEBM Levels of Evidence. (2022). Centre for Evidence-Based Medicine (CEBM), University of Oxford [Internet]. [cited 2022 Mar 10]. Available from: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence
9. Grading Guide | UpToDate | Wolters Kluwer [Internet]. [cited 2022 Mar 10]. Available from: https://www.wolterskluwer.com/en/solutions/uptodate/policies-legal/grading-guide
10. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ [Internet]. 2008 Apr 24 [cited 2022 Mar 10];336(7650):924–6. Available from: https://www.bmj.com/content/336/7650/924
11. Kim KW, Lee J, Choi SH, Huh J, Park SH. Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part I. General Guidance and Tips. Korean J Radiol [Internet]. 2015 Nov 1 [cited 2022 Mar 13];16(6):1175. Available from: /pmc/articles/PMC4644738/
12. West SL, Gartlehner G, Mansfield AJ, Poole C, Tant E, Lenfestey N, et al. Comparative Effectiveness Review Methods: Clinical Heterogeneity. Comp Eff Rev Methods Clin Heterog [Internet]. 2010 [cited 2022 Mar 13]; Available from: https://www.ncbi.nlm.nih.gov/books/NBK53310/
13. Wallace BC, Dahabreh IJ, Trikalinos TA, Lau J, Trow P, Schmid CH. Closing the Gap between Methodologists and End-Users: R as a Computational Back-End. J Stat Softw [Internet]. 2012 Jun 30 [cited 2022 Mar 11];49:1–15. Available from: https://www.jstatsoft.org/index.php/jss/article/view/v049i05
14. Freeman SC, Kerby CR, Patel A, Cooper NJ, Quinn T, Sutton AJ. Development of an interactive web-based tool to conduct and interrogate meta-analysis of diagnostic test accuracy studies: MetaDTA. BMC Med Res Methodol [Internet]. 2019 Apr 18 [cited 2022 Mar 11];19(1):1–11. Available from: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-019-0724-x

15. Okan Akçam M, Takada K. Fuzzy modelling for selecting headgear types. Eur J Orthod [Internet]. 2002 [cited 2022 Mar 12];24(1):99–106. Available from: https://pubmed.ncbi.nlm.nih.gov/11887385/
16. Choi, H, Il, Jung, S.K., Baek, S.H., Lim, W.H., Ahn, S.J., Yang. I.H. (1986). Artificial Intelligent Model with Neural Network Machine Learning for the Diagnosis of Orthognathic Surgery. J Craniofac Surg [Internet]. 2019 Oct 1 [cited 2022 Mar 12];30(7); 9. Available from: https://journals.lww.com/jcraniofacialsurgery/Fulltext/2019/10000/Artificial_Intelligent_Model_With_Neural_Network.16.aspx
17. Jung, S. K., & Kim, T. W. (2016). New approach for the diagnosis of extractions with neural network machine learning. *American Journal of Orthodontics and Dentofacial Orthopedics*, *149*(1), 127-133.
18. Kim, Y. H., Park, J. B., Chang, M. S., Ryu, J. J., Lim, W. H., & Jung, S. K. (2021). Influence of the Depth of the Convolutional Neural Networks on an Artificial Intelligence Model for Diagnosis of Orthognathic Surgery. *Journal of Personalized Medicine*, *11*(5), 356.
19. Li, P., Kong, D., Tang, T., Su, D., Yang, P., Wang, H., ... & Liu, Y. (2019). Orthodontic treatment planning based on artificial neural networks. *Scientific reports*, *9*(1), 1-9.
20. Lin, G., Kim, P. J., Baek, S. H., Kim, H. G., Kim, S. W., & Chung, J. H. (2021). Early prediction of the need for orthognathic surgery in patients with repaired unilateral cleft lip and palate using machine learning and longitudinal lateral cephalometric analysis data. *Journal of Craniofacial Surgery*, *32*(2), 616-620.
21. Shin, W., Yeom, H. G., Lee, G. H., Yun, J. P., Jeong, S. H., Lee, J. H., ... & Kim, B. C. (2021). Deep learning based prediction of necessity for orthognathic surgery of skeletal malocclusion using cephalogram in Korean individuals. *BMC Oral Health*, *21*(1), 1-7.
22. Suhail, Y., Upadhyay, M., & Chhibber, A. (2020). Machine learning for the diagnosis of orthodontic extractions: a computational analysis using ensemble learning. *Bioengineering*, *7*(2), 55.
23. Xie, X., Wang, L., & Wang, A. (2010). Artificial neural network modeling for deciding if extractions are necessary prior to orthodontic treatment. *The Angle Orthodontist*, *80*(2), 262-266.
24. Thanathornwong, B. (2018). Bayesian-based decision support system for assessing the needs for orthodontic treatment. *Healthcare informatics research*, *24*(1), 22-28.
25. Khanagar, S. B., Al-Ehaideb, A., Vishwanathaiah, S., Maganur, P. C., Patil, S., Naik, S., ... & Sarode, S. S. (2021). Scope and performance of artificial intelligence technology in orthodontic diagnosis, treatment planning, and clinical decision-making-A systematic review. *Journal of dental sciences*, *16*(1), 482-492.
26. Nieri, M., Crescini, A., Rotundo, R., Baccetti, T., Cortellini, P., & Prato, G. P. P. (2010). Factors affecting the clinical approach to impacted maxillary canines: A Bayesian network analysis. *American journal of orthodontics and dentofacial orthopedics*, *137*(6), 755-762.
27. Mary, A. V., Mahendra, J., John, J., Moses, J., Ebenezar, A. R., & Kesavan, R. (2017). Assessing quality of life using the oral health impact profile (OHIP-14) in subjects with and without orthodontic treatment need in chennai, tamil nadu, India. *Journal of clinical and diagnostic research: JCDR*, *11*(8), ZC78.
28. Luke, L. S., Atchison, K. A., & White, S. C. (1998). Consistency of patient classification in orthodontic diagnosis and treatment planning. *The Angle Orthodontist*, *68*(6), 513-520.