Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(5), 1731-1744 | ISSN: 2251-6719

Diagnostic Accuracy of Artificial Intelligence to Predict the Need for Orthodontic Extraction Versus Non-Extraction- A Systematic Review and Meta-Analysis

Dr. Romilkumar Shah[1], Dr. Yashraj Kharade[2], Dr. Prishita Mehta[3], Dr. Nancy Agrawal[4], Dr. Anjali Ganatra[5]

[1](Department of Orthodontics and Dentofacial Orthopedics, K.M. Shah Dental College and Hospital/ Sumandeep Vidyapeeth Deemed to be University, India)

[2](Department of Orthodontics and Dentofacial Orthopedics, K.M. Shah Dental College and Hospital/ Sumandeep Vidyapeeth Deemed to be University, India)

[3](Department of Orthodontics and Dentofacial Orthopedics, K.M. Shah Dental College and Hospital/ Sumandeep Vidyapeeth Deemed to be University, India)

[4](Department of Orthodontics and Dentofacial Orthopedics, K.M. Shah Dental College and Hospital/ Sumandeep Vidyapeeth Deemed to be University, India)

[5](Department of Orthodontics and Dentofacial Orthopedics, K.M. Shah Dental College and Hospital/ Sumandeep Vidyapeeth Deemed to be University, India)

(Received: 24 July 2025 Revised: 28 August 2025 Accepted: 26 September 2025)

KEYWORDS

Fixed orthodontics, automated artificial intelligence, Machine Erudition, Diagnostic accuracy, decision making, extraction diagnosis.

ABSTRACT:

Objective:
To evaluate the diagnostic accuracy of artificial intelligence (AI) models in predicting the necessity for orthodontic extractions versus non-extractions.

Material and Methods – A systematic review was conducted following PRISMA-DTA guidelines, with studies selected based on the PIRD (Population, Index test, Reference standard, Diagnosis) framework. The inclusion criteria focused on diagnostic accuracy studies that used AI models to predict orthodontic extractions. Searches were performed across PubMed, DOAJ, EBSCO, K-hub, and Google Scholar for articles published from January 2000 to December 2023. 12 studies met the criteria, with a pooled meta-analysis involving four studies. Data were extracted and assessed for quality using QUADAS-2. Diagnostic accuracy metrics, including sensitivity, specificity, and receiver operating characteristic (ROC) curves, were analyzed using MetaDiSc version 1.4.

Results –The meta-analysis of 4 studies revealed a pooled sensitivity of 0.73 (95% CI: 0.70–0.76) and specificity of 0.82 (95% CI: 0.80–0.84), The ROC curve yielded an area under the curve (AUC) of 0.8813, Subgroup analysis highlighted that multi-layer perceptron (MLP) models had a sensitivity of 0.83 (95% CI: 0.79–0.87).

Conclusion - PEEK Retainers demonstrated superior Survival Rates and comparable Stability and Periodontal Health outcomes compared to Conventional Retainers. Although both materials performed acceptably, PEEK may offer enhanced durability, making it a viable alternative for long-term retention.

1. INTRODUCTION

Improvement of our knowledge of accuracy of diagnostic test can be done by systematic reviews.[1] Comprehensive systematic analyses of research on diagnostic accuracy are often characterized by markedly heterogeneous results originating from differences within the design layout and demeanor of included studies.[2]

However, orthodontic diagnosis is frequently exceedingly challenging and impacted by the subjective scrutiny of the criteria that are examined. Because of this, methods based on artificial expert system (AESs) have been suggested as a reliable aid for orthodontic diagnosis.[3]

Despite the popularity of extracting teeth in orthodontic practice, there aren't any objective standards to be used by orthodontists to decide whether to extract or not to extract teeth.[4] (ANN) is a computer model designed to replicate how the human brain processes signals through neurons and their connections. By understanding biological neural networks, ANNs can analyze complex data to provide predictions for diagnosis, treatment, and outcomes without pre-defining rules.[5]

Neurons connect the input to the output in (ML) using neural networks, in machine learning with neural networks, neurons connect inputs to outputs via synapses that use weighted values. Repeated learning adjusting weightiness to enhance fit, but excessive iterations can lead to overfitting.[6] The (ANN) has the advantage of excavating features from massive medical data.[7] The (ANN) not only limit variability in decision-making in orthodontics but also lessen the negative consequences of improperly recommended tooth extraction.[8]

Machine learning is a rapidly evolving field with application in every aspect of our lives.[9] In orthodontics, ANNs help determine tooth extraction needs and patterns. In healthcare, AutoML systems can reduce costs, enhance health outcomes, and advance clinical research. Their ease of use allows non-experts, such as orthodontists, to develop and utilize AI systems independently[10].

(ML) models fall into supervised and unsupervised categories. Supervised type of learning utilizes labelled data input and output, prepared by scientists before training. This method involves human oversight, allowing the model to classify and predict new data based on learned relationships.[11]

Unsupervised (ML) trains models on unlabeled, raw data for identifying patterns, trends, or group similar data into predefined clusters. This approach, requiring minimal human intervention, is often used for initial data exploration and understanding.[11]

This comprehensive systematic review and meta-analysis aimed to evaluate the diagnostic accuracy of artificial intelligence in predicting the necessity of orthodontic extractions. The objective was to critically assess the effectiveness of AI-based technologies in determining whether orthodontic cases require extraction or can be managed with non-extraction treatment approaches.

2. MATERIAL AND METHODS

Comprehensive systematic research of literature and meta-analysis was performed. This study followed the PRISMA-DTA guidelines for assessing the accuracy of diagnostic research, the Cochrane Handbook for Systematic Reviews of Interventions (version 5.1.0), and the JBI Reviewer's Manual (Fourth Edition), and was registered at PROSPERO with registration code CRD42024501466. The following focused question in the Population, Index test, Reference standard, and Diagnostic index (PIRD) format was proposed: "Is there a difference in the accuracy of diagnostic information of artificial intelligence to envisage the necessity for orthodontic extraction versus non-extraction in decision making?

Study Design: Systematic Review and Meta Analysis

Eligibility criteria:

[A] Inclusion criteria:

For diagnostic accuracy studies, PIRD is recommended (instead of PICOS). Population, Index test, Reference standard, and Diagnostic of interest are the acronyms for PIRD.

a. Population –

Studies with patients employing either buccal or lingual orthodontic fixed appliances in a series of treatments involving comprehensive orthodontic conduct in the permanent dentition, having complete standardized orthodontic pretreatment records such as patient photographs, intra-oral photographs, X-rays, models, cephalometric analysis, etc.

b. Index test –

Studies using AI-based models for decision-making on tooth extraction in orthodontic treatment.

c. Reference standard –

Studies using pretreatment records of orthodontics such as patient photographs, intra- and extraoral photographs, dental models, X-rays, and lateral cephalometry evaluated by experts in orthodontics.

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(5), 1731-1744 | ISSN: 2251-6719

d. Diagnostic interest –

Studies providing information about diagnostic accuracy including sensitivity, specificity, accuracy, PPV, and NPV determined using different methods, regardless of the methods of quantifying the outcomes.

e. Study design –

i. Studies published in any language where English translation is possible.

ii. Studies published between 01-01-2000 to 31-12-2023.

iii. Diagnostic precision studies.

iv. Studies with full-text articles.

v. Studies providing a numeric value to calculate at least any outcome measures mentioned

earlier.

vi. Studies were included only if the results were compared to a control group or to a standard benchmark.

[B] Exclusion criteria:

a. Partially accessible research papers in the repository.

b. Studies involving participants with congenital abnormalities, incomplete records, orthognathic surgical therapy, or missing one or more teeth at baseline (apart from the third molars).

c. Studies with AI not investigating orthodontic tooth extraction decision-making.
d. Single intervention trials without a control group.

Search strategy:

Studies were selected based on the PIRD inclusion criteria. Titles and abstracts which detected eligible studies were checked by two reviewers. Any enquiries were discussed with a third reviewer. Medical Subject Headings (MeSH) terms, keywords, and other free terms combined with Boolean operators (OR, AND) were used for searching. All search platforms employed the same keywords, complying with the syntactic guidelines.

Table 1: The PICOS tool plus search approach

Search strategy	
Focused Question	Is there a difference in the accuracy of diagnostic information of autonetic intelligence to envisage the need for orthodontic extraction versus non-extraction in decision making?
Search strategy	
Population (#1)	(((Human teeth [Text Word]) OR "tooth"[MeSH Terms] OR teeth [Text Word]) OR permanent teeth [Text Word] OR permanent dentition [Text Word] OR fixed orthodontics [Text Word] OR [Text Word] buccal orthodontics [Text Word] OR lingual orthodontics [Text Word])))
Index test (#2)	(‘(ML)’ [MeSH] OR ‘(ML)’ [Text Word] OR Multilayer perception [Text Word] OR Machine Intelligence [Text Word] OR ANN [Text Word] OR automated artificial intelligence [Text Word] OR ensemble learning [Text Word])
Reference standard (#3)	((pretreatment records [Text Word] OR photographs [Text Word] OR models [Text Word] OR cephalometric analysis [Text Word]))
Diagnostic interest (#4)	(((Diagnostic accuracy [Text Word] OR Sensitivity [Text Word] OR Accuracy [Text Word] OR Specificity [Text Word]))) AND ‘orthodontic treatment needs’ [Text Word] OR decision making [Text Word] OR ‘extraction diagnosis’[Text Word] OR ‘orthodontic treatment planning’ [Text Word])))
Study design (#5)	(Clinical trial [MeSH] OR clinical study [Text Word] OR randomized controlled trials [Text Word] OR randomized control clinical trial [MeSH] MeSH, non-randomized control trials, quasi-experimental studies, before-and-after study design, cohort studies, in vivo research, and cross-sectional studies, among other possibilities)
Search Combination	#1 AND #2 AND #3 AND #4 AND #5
Database search	
Linguistic	No restraint (Articles in English language or other language where English translation is possible.)

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(5), 1731-1744 | ISSN: 2251-6719

Electronic Databases	PubMed/MEDLINE, Cochrane Central Register of Controlled Trials, Web of Science, Open grey, Google scholar

Primary outcomes measured were both specificity and sensitivity of the AI-based models as compared to pre-treatment records. To do a meta-analysis, stick to the Chosen Reporting Items for Systematic Reviews & Meta-Analyses (PRISMA). The electronic data resources consulted were PubMed, DOAJ, EBSCO, K-hub, and Google Scholar search engine with controlled vocabulary and free text terms.Articles published from 01/01/2010 until 31/12/2023 were searched, without any restriction concerning language.

Selection of studies:

The title and the abstract of each study were reviewed and critically assessed by two independent reviewers. The methods used to apply the selection criteria were the following:

i. Integration of the searched outcomes to delete duplicate entries.

ii. Review of abstracts and titles to eliminate pointless articles.

iii. Retrieval of pertinent articles' entire texts.

iv. Binding and compiling several research portions into one.

v. Reading through the articles in their entirety to confirm how closely the studies adhered to

the qualifying requirements.

vi. Establishing connection with researchers, if necessary, to clarify the study's eligibility.

vii. Deciding about the study's inclusion and data gathering.

Data extraction:

Extraction of data from studies was done separately by two reviewers, resolving disagreements through discussion, using a verification list with key items.

Authors, Year and Title of study

Country

Sample size

Age group and Gender of participants

AI model used

Reference standard used

Time period between two examinations

Number of examiners used for assessment

Outcomes

Results and other items

Quantitative sensitivity and specificity data were gathered from each study. Using these, true positive, true negative, false positive, and false negative values were manually calculated where not provided. Corresponding authors were emailed for supplementary information if needed.

a. False positive = (1 − specificity) × (1 − diseased cases / total sample)
b. True negative = specificity × (1 − diseased cases / total sample)
c. True positive = sensitivity × diseased cases / total sample
d. False negative = (1 − sensitivity) × diseased cases / total sample

Publication details, study characteristics, participant information, settings, interventions, comparators, outcome measures, study design, statistical analysis, results, plus additional pertinent data were meticulously extracted and recorded in Excel sheets for primary outcomes.

Evaluations of the potential for bias and quality:

The QUADAS-2 tool was employed to evaluate the operational excellence of selected studies, following guidelines from Cochrane, the UK NIH and Clinical Excellence, and the Agency for Healthcare Research and Quality. This instrument assesses applicability issues and bias risk in four domains: stream and control, reference standard, index test, and patient selection. Each domain is examined for issues related to applicability and bias risk. Quality was gauged using Review Manager version 5.4.

Data synthesis and analysis:

A meta-analysis was directed on studies with similar outcomes using MetaDiSc version 1.4, examining sensitivity and specificity values. Contingency tables with true positive, true negative, false positive, and false

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(5), 1731-1744 | ISSN: 2251-6719

negative results were created. A bivariate diagnostic meta-analysis provided collective evaluations of sensitivity, specificity, likelihood ratios, diagnostic odds ratios, and 95% CI, accounting for between-study variability.

Heterogeneity:

The heterogeneity between all the studies was calculated using the I^2 statistic and Chi-square test. An I^2 value of 25% was considered low, 50% moderate, and 75% high. By visually examining the summary receiver operation characteristic curves and determining which studies fell outside the 95% confidence ellipse, the foundations of heterogeneity were investigated. An I^2 value > 50% or p-value < 0.05 was considered substantial heterogeneity, and a random effect model was selected owing to the significant heterogeneity across the studies and applied in all forest plots.

3. RESULTS

Study Characteristics

Search of Literature:

The search result of initial electronic database showed a over-all of 239 titles (95 titles in PubMed/MEDLINE and Cochrane library and 144 titles in google scholar, along with no additional papers were discovered while searching the orientation slants of the chosen studies by hand), removal of exactly similar titles, 68 titles were left. After going over the abstracts and titles of these 239 documents, 171 among them were discarded during the preliminary screening. Once the reviewers glanced through and discussed the articles, 25 were picked for full-text evaluation. Four research investigations were counted in the meta-analysis and 12 investigations were contained within in the qualitative analysis upon pre-screening procedures and application of the eligibility criteria

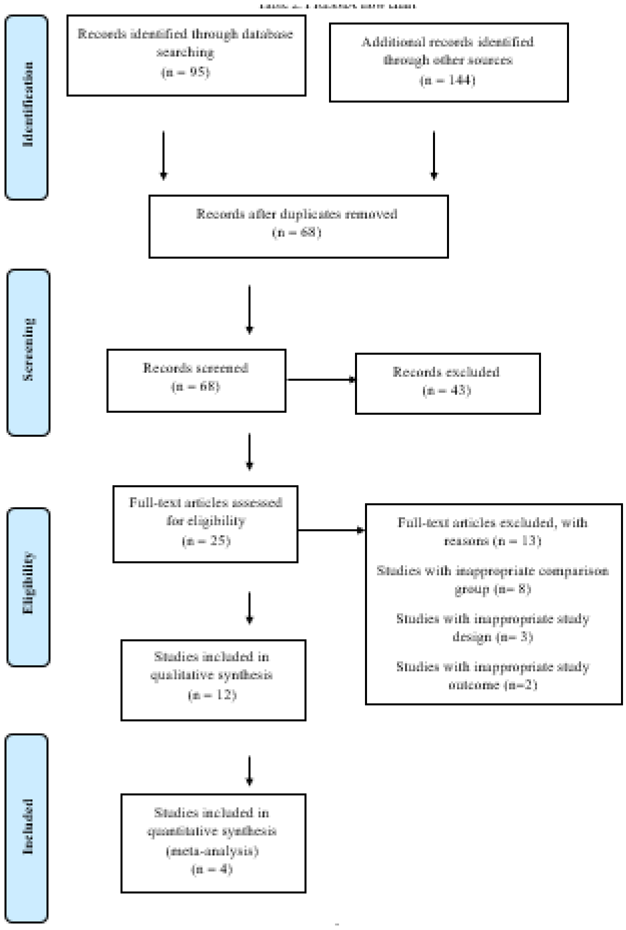


Table 2: PRISMA flow chart

The systematic review encompassed twelve studies, the general features of which are

shown in Table 3: Features of the studies that were incorporated

Table 3: Features of the studies that were incorporated

Author name and year	Study design	Total Sample sets n (100%)	Age/ Gender	Training sets n(%)	Validity sets n(%)	Test sets n(%)	Sample characteristics extraction cases	Sample characteristics non-extraction cases	Experts Training/n /experience	Diagnostic tools	AI type (algorithm)	Extraction diagnosis model	Authors Conclusion

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(5), 1731-1744 | ISSN: 2251-6719

Martina 2006	cross-sectional	48	9-26 / NA	N/A	N/A	N/A	N/A	N/A	N/A	orthodontic casts & radiographic measurements, clinical examination	neural network	N/A	ANN using cephalometric and cast measurements aid orthodontic extraction decisions.
Takada 2009	cross-sectional	188	12-36 years / Females	N/A	N/A	N/A	188 patient records	N/A	3	5 cephlometric records, 22 clinical records	(ML) (Feature vectors).	N/A	A mathematical model simulates expert tooth extraction decisions with 90.4% accuracy.
Xie 2010	cross-sectional	120 - extraction 80 - non-ex	11-15 years / NA	180	N/A	20	120 cases	80 cases	N/A	5 cast measurements, 13 cephalometric landmarks, 5 soft tissue landmarks	ANN model	decision making BP ANN model	The ANN achieved 80% accuracy in treatment decisions for malocclusion in 11-15-

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(5), 1731-1744 | ISSN: 2251-6719

													year-olds.
Jung 2016	cross-sectional	156	NA / 94 M, 62 F	64	32	60	94	62	N/A	cast measurements, clinical records, cephalometric landmarks	(ML)	ANN model	AI expert systems with neural networks could enhance orthodontic performance through optimal data selection and modeling.
Li 2019	cross-sectional	302	N/A	60%	20%	20%	222 (73.5%)	80 (26.5%)	N/A	cast measurements, clinical records, cephalometric landmarks	(ANN)	Multilayer perceptron (MLP)	The method using ANNaids less-experienced orthodontists in treatment planning.
Suhail 2020	retrospective	287	N/A	N/A	N/A	N/A	N/A	N/A	5 / 9 years experience	14 diagnostic outcomes	(ML) - random forest mode	N/A	Our feature set and algorithm predict extract

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(5), 1731-1744 | ISSN: 2251-6719

											ls, decision trees		ions with expert-like accuracy.
Etemad 2021	cross-sectional	838	NA / 341 M, 497 F	90 %	N/A	10 %	208	630	19 orthodontists	117 Cephalometric and clinical components	(ML)	arbitrary forest (AF), MLP model	Incongruent data patterns require future AI improvements for clinical decision-making.
Real 2022	cross-sectional	214	NA / 94 M, 120 F	N/A	N/A	N/A	81	133	2 / 30 years experience	cephalometric data	automated (ML) software (Auto-WEKA). 3 algorithms used - Bagging, Random committee, Multilayer	models set in three settings and five-time limit sets.	Automated (ML) generates accurate orthodontic extraction models using cephalometric data.

											perceptron		
Etemad 2023	cross-sectional	1135	mean: 18.37 +- 10.69 / 472 M, 663 F	942	N/A	193	301	834	18	9 clinical, 11 cephalometric parameters	(ML)	arbitrary forest (AF)	AI systems enhance extraction decisions, with crowding as the key factor.
Leavit 2023	retrospective	366	126/240	70%	N/A	30%	Patients who had been treatment planned for orthodontic comprehensive care with premolar extractions were included in the study	N/A	30	N/A	(ML)	LR from linear models, Random Forest (RF) from tree-based models, and Provide Vector Machine Support (SVM)	Supervised ML techniques accurately predicted U/L4s and U4s extractions but struggled with others. Molar relationship, mandibular crowding, and overjet were key

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(5), 1731-1744 | ISSN: 2251-6719

													indicators.
Mason 2023	cross-sectional	393	143 M, 250 F	70 %	N/A	30 %	193	200	1	2 demographic, 4 clinical and 50 cephalometric features	(ML)	logistic regression (LR), random forest (RF), Vector Machine Support (SVM), NEURAL NETWORK (NN)	ML models accurately predict extractions across diverse populations, emphasizing crowding, sagittal, and vertical features.

Studies were carried out across the globe. – Italy[3], Japan[4], China[5,7], Korea[6,14], USA[8,9,12,13], Chile[10]. All the included studies showed cross-sectional study design. In this research, evaluation was carried out using clinical records from 5763 patients in total.

Standard of reference

Cast measurements, soft tissue landmarks, clinical chronicles, and cephalometric records served as the reference standards for every included investigation.Most studies employed seasoned examiners with at least 8–10 years of clinical practice and an orthodontic specialization. Leavit 2023[12] employed a panel of thirty specialized examiners to evaluate the clinical records.

Index test

Various algorithms for AI were implemented in these investigations – neural networks[3,5,7,] (ML)[4,6,8,9,11-13,] deep learning models14, etc. Under (ML), different systems such as random forest, logistic regression, support vector machine, etc. were used.From the deductions of all studies, Additionally, it was pointed out that the decision-making properties of AI based models show precision almost equal to that of the clinical examiners.

Bias applicability risk:

The QUADAS-2 analyser assessed the Caliber of investigation on diagnostic exactness based on bias and applicability concerns across four domains: patient selection, movement and execution, standard of reference, and indexing test. Most studies were of acceptable to moderate quality, but three had high risk, particularly linked to the indexing test. Bias is uncertain due to convenience sampling and uneven sample sizes for extraction and non-extraction cases, raising concerns about patient selection applicability.

Figure 1: Bias graph risk

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(5), 1731-1744 | ISSN: 2251-6719

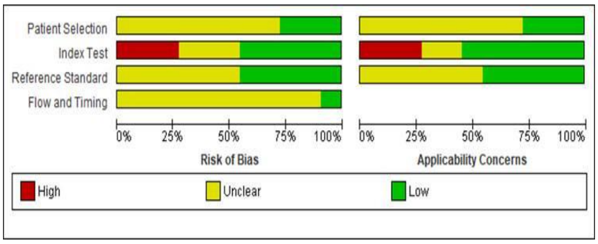
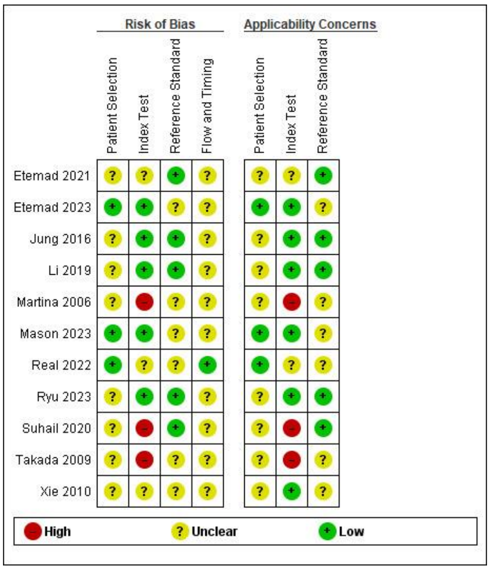


Figure 2: Summary of bias risk



Meta-analysis:

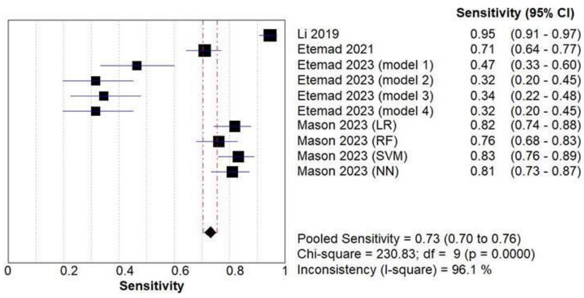
A. Pooled analysis

Four studies[7,9,11,13] was incorporated into the combined diagnostic precision, regardless of AI model type. Etemad 2023 and Mason 2023, using four different AI models, were analyzed separately.

1. **Sensitivity**

   Pooled sensitivity obtained was 0.73[0.70, 0.76] (true positives). Overall, the upshots were statistically significant ($p<0.05$) with 96.1% heterogeneity.

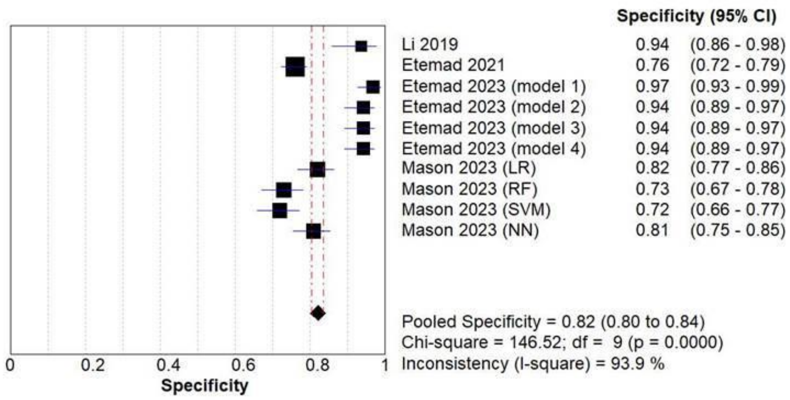
Figure 3: Pooled sensitivity



Specificity

Pooled specificity obtained was 0.82[0.80, 0.84] (true negative). Overall, the outcomes were statistically significant ($p<0.05$) with 93.9% heterogeneity.

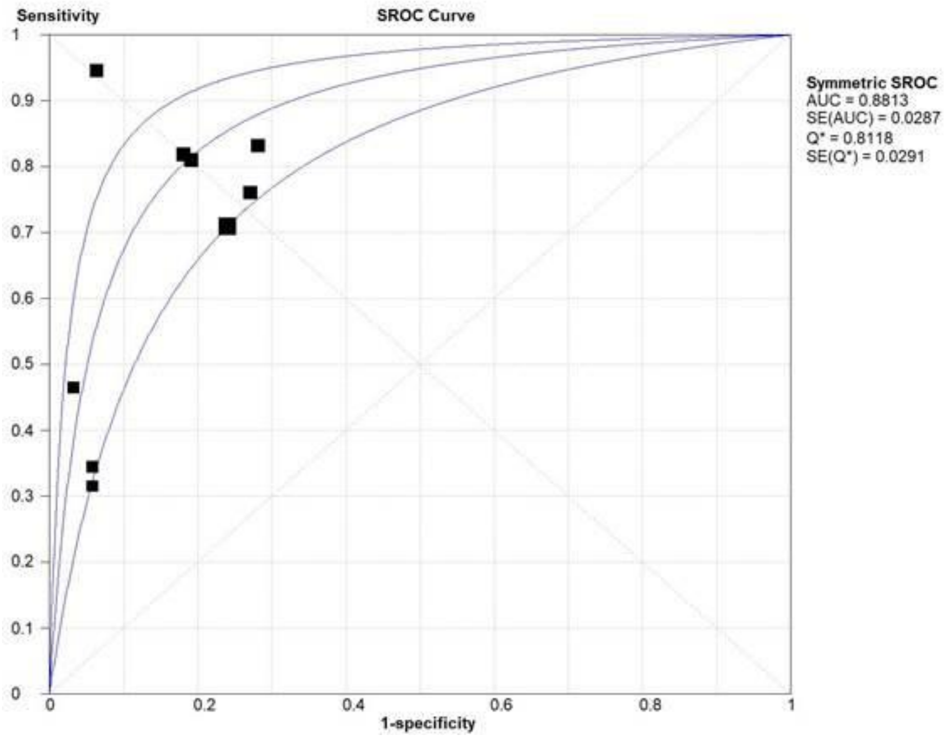
Figure 4: Pooled specificity



2. **Receiver Operating Characteristics:**

Characteristics of the receiver operation curve showed that 88.13% of data was beneath the curvature (Area UC = 0.8813) which indicated good performance of AI models in predicting orthodontic extraction.

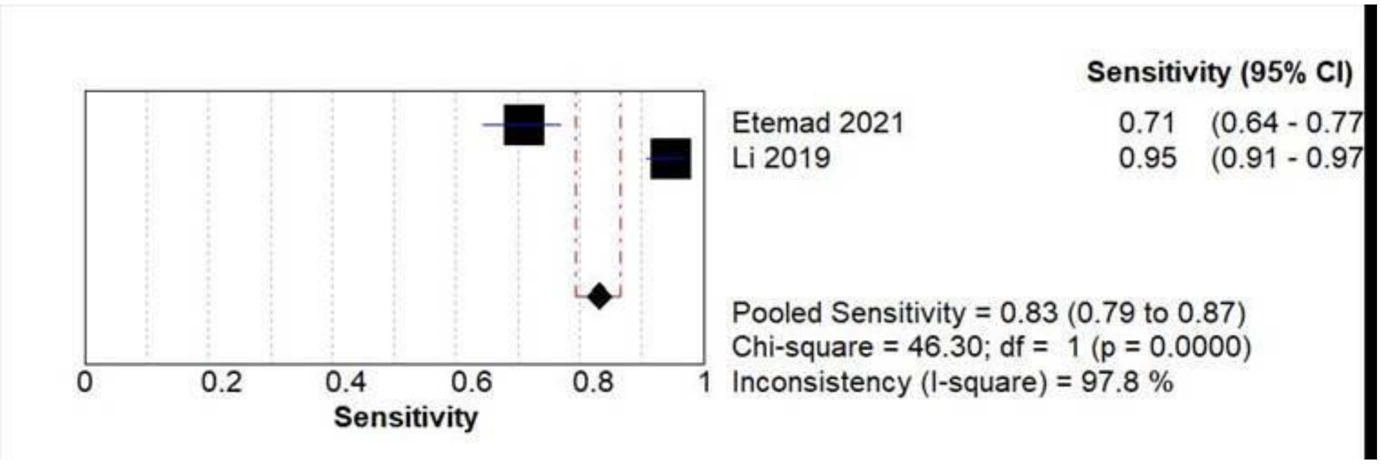
Figure 5: Characteristics of the receiver operation



Subgroup analysis:

Due to significant heterogeneity among studies, a subgroup scrutiny performed based on AI model type. For multi layered perceptron (MLP) model, evaluated in studies by Etemad 2021 and Li 2019, sensitivity was 0.83 [0.79, 0.87], indicating 83% precision in recognizing orthodontic extraction needs. Results were statistically significant ($p<0.05$) with 97.8% heterogeneousness, and the MLP model's sensitivity exceeded the collective sensitivity.

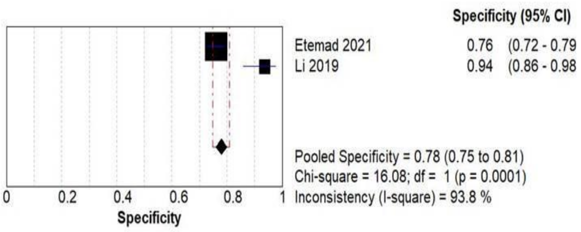
Figure 6: Sensitivity of MLP



Specificity obtained was 0.82[0.80, 0.84] (true negative). Overall, the results were statistically significant ($p<0.05$) with 93.8% heterogeneity. The exactness of The exactness of MLP model was beneath the collective sensitivity.

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(5), 1731-1744 | ISSN: 2251-6719

Figure 7: Specificity of MLP



DISCUSSION

Recent interest has surged in Utilizing A. intelligence (AI) so as to improve planning of orthodontic procedures. This meta-analysis and systematic review appraise AI models' specificity of diagnosis for orthodontic extractions, analyzing results from twelve diverse studies.

The search of literature, detailed in the appraisal, identified twelve studies evaluating various AI Programs. Studies, which included varied demographics of patients, provided insights into AI's effectiveness in orthodontic decision-making. Most studies were cross-sectional (10 out of 12), using clinical examinations and cephalometric data[3,4,5,6,7,9,10,11,13,14,] while two were retrospective.[8,12] Different AI methods, including neural networks[3,5,7,] (ML)[4,6,8,9,10,11,12,13,] and deep learning models[14], were utilized in search operation. This diversity shows AI's adaptability in orthodontics and emphasizes the thoroughness of advanced diagnostic tools. Overall, research demonstrates that AI programs significantly enhance orthodontic treatment preparation with high diagnostic accuracy, comparable to expert clinical examiners. AI models excel at assessing factors like sagittal features and dental crowding impacting extraction decisions. The meta-analysis offers valuable insights, including combined estimations of specificity and sensitivity and ROC curves.

Four studies[7,9,11,13] was incorporated into the appraisal of pooled diagnostic accuracy. Two studies Etemad 2023 and Mason 2023 used four different AI models for appraisal of extraction hence they were included separately in the analysis. Collective sensitivity of 0.73 [0.70, 0.76] indicates that AI replicas correctly identified 73% of cases needing orthodontic extraction. With 96.1% heterogeneity, results were statistically significant ($p<0.05$). The pooled specificity was 0.82 [0.80, 0.84], showing 82% accuracy in detecting cases not needing extraction, with 93.9% heterogeneity and overall statistical significance ($p<0.05$). The receiver operating distinctive curve, with the area inside curvature of 0.8813, demonstrated AI simulations performed well in predicting orthodontic extractions. Due to significant heterogeneity, a subgroup analysis by AI model type was conducted. Subgroup analysis was carried out on multi-layer perceptron model (MLP) weighed by two of studies Li 20197 and Etemad 20219.

The MLP model showed a perceptiveness of 0.83 [0.79, 0.87], correctly identifying 83% of cases needing orthodontic extraction, with results statistically significant ($p<0.05$) despite 97.8% heterogeneity. Its specificity was 0.82 [0.80, 0.84], accurately identifying 78% of cases not needing extraction, with 93.8% heterogeneity and significant results ($p<0.05$). Despite these promising results, variability in study quality and sample sizes poses limitations, highlighting requirement for larger, well-designed studies to improve robustness and generalizability.

3. CONCLUSION

In conclusion, our meta-analysis and comprehensive systematic review provide compelling evidence supporting the usefulness of AI-based algorithms in planning of orthodontic procedures. The findings underscore the AI's Potential to augment clinical decision-making, offering a pathway towards exact and personalized orthodontic interventions. Moving forward, further research endeavors should emphasis on addressing methodological limitations and validating the clinical applicability of AI simulations in diverse patient cohorts.

REFERENCES

1. McInnes MDF, Moher D, Thombs BD, et al. Preferred Reporting Items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA statement: The PRISMA-DTA statement. *JAMA*. 2018;319(4):388-396. doi:10.1001/jama.2017.19163
2. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2 Group*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529-536.
3. Martina R, Teti R, 'addona D, Iodice DO. Neural Network Based System for Decision Making Support in Orthodontic Extractions.

In: Intelligent Production Machines and Systems. Elsevier Science Ltd; 2006.

4. Takada K, Yagi M, Horiguchi E. Computational Formulation of Orthodontic Tooth-Extraction Decisions: Part I: to extract or not to extract. *Angle Orthod*. 2009;79:885-891.
5. Xie X, Wang L, Wang A. Artificial neural network modeling for deciding if extractions are necessary prior to orthodontic treatment. *Angle Orthod*. 2010;80(2):262-266. doi:10.2319/111608-588.1
6. Jung SK, Kim TW. New approach for the diagnosis of extractions with neural network machine learning. *Am J Orthod Dentofacial Orthop*. 2016;149(1):127-133. doi:10.1016/j.ajodo.2015.07.030
7. Li P, Kong D, Tang T, et al. Orthodontic treatment planning based on artificial neural networks. *Sci Rep*. 2019;9(1):2037. doi:10.1038/s41598-018-38439-w
8. Suhail Y, Upadhyay M, Chhibber A, Kshitiz. Machine learning for the diagnosis of orthodontic extractions: A computational analysis using ensemble learning. *Bioengineering (Basel)*. 2020;7(2):55. doi:10.3390/bioengineering7020055
9. Etemad L, Wu TH, Heiner P, et al. Machine learning from clinical data sets of a contemporary decision for orthodontic tooth extraction. *Orthod Craniofac Res*. 2021;24 Suppl 2(S2):193-200. doi:10.1111/ocr.12502
10. Real AD, Real OD, Sardina S, Oyonarte R. Use of automated artificial intelligence to predict the need for orthodontic extractions. *Korean J Orthod*. 2022;52(2):102-111. doi:10.4041/kjod.2022.52.2.102
11. Etemad L. *Machine Learning with Cross-Institutional Prediction for Orthodontic Tooth Extraction* (Master’s Thesis).
12. Leavitt L, Volovic J, Steinhauer L, et al. Can we predict orthodontic extraction patterns by using machine learning? *Orthod Craniofac Res*. 2023;26(4):552-559. doi:10.1111/ocr.12641
13. Mason T, Kelly KM, Eckert G, Dean JA, Dundar MM, Turkkahraman H. A machine learning model for orthodontic extraction/non-extraction decision in a racially and ethnically diverse patient population. *Int Orthod*. 2023;21(3):100759. doi:10.1016/j.ortho.2023.100759
14. Ryu J, Kim YH, Kim TW, Jung SK. Evaluation of artificial intelligence model for crowding categorization and extraction diagnosis using intraoral photographs. *Sci Rep*. 2023;13(1):5177. doi:10.1038/s41598-023-32514-7