



UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE – CCBS
DEPARTAMENTO DE ODONTOLOGIA – DOD

ISADORA MARIA BATISTA DA SILVA MOTA

**A MARCAÇÃO CEFALOMÉTRICA AUTOMÁTICA COM INTELIGÊNCIA
ARTIFICIAL É CONFIÁVEL? REVISÃO SISTEMÁTICA COM METANÁLISE DAS
MEDIDAS CEFALOMÉTRICAS SNA, SNB, ANB E WITS**

ARACAJU

2022

ISADORA MARIA BATISTA DA SILVA MOTA

**A MARCAÇÃO CEFALOMÉTRICA AUTOMÁTICA COM INTELIGÊNCIA
ARTIFICIAL É CONFIÁVEL? REVISÃO SISTEMÁTICA COM METANÁLISE DAS
MEDIDAS CEFALOMÉTRICAS SNA, SNB, ANB E WITS**

Monografia apresentada ao Departamento de Odontologia como pré-requisito à conclusão do Curso de Odontologia da Universidade Federal de Sergipe para obtenção do grau de Cirurgiã-dentista.

Orientador: Prof. Dr. Wilton Mitsunari Takeshita

Coorientador: Prof. Me. Marcos Antônio Lima dos Santos

ARACAJU

2022

ISADORA MARIA BATISTA DA SILVA MOTA

**A MARCAÇÃO CEFALOMÉTRICA AUTOMÁTICA COM INTELIGÊNCIA
ARTIFICIAL É CONFIÁVEL? REVISÃO SISTEMÁTICA COM METANÁLISE DAS
MEDIDAS CEFALOMÉTRICAS SNA, SNB, ANB E WITS**

Aracaju, ____/____/____.

Monografia aprovada como requisito parcial à conclusão do Curso de Odontologia da Universidade Federal de Sergipe para obtenção do grau de cirurgião-dentista.

Prof. Dr. Wilton Mitsunari Takeshita - Orientador
Universidade Federal de Sergipe

Prof. Me. Marcos Antônio Lima dos Santos - Coorientador

Prof.^(a) Dr.(a). - 1º Examinador
Universidade Federal de Sergipe

Prof.^(a) Dr.(a). - 2º Examinador
Universidade Federal de Sergipe

AGRADECIMENTOS

Agradeço primeiramente a Deus, que me deu forças para vencer todas as etapas e que me abençoou para que eu chegasse até aqui. Obrigada, senhor, por estar ao meu lado em todos os momentos. “Não fui eu que ordenei a você? Seja forte e corajoso, pois o senhor, o seu Deus, estará com você por onde você andar.” Josué 1:9.

Aos meus pais, Cristiany e Ronivon, que me deram todo o apoio possível e que sempre lutaram pela minha educação. Agradeço também ao meu irmão, Victor. Amo vocês.

Ao meu amor, Paulo, agradeço por estar ao meu lado em grande parte desta caminhada. Amo você.

À toda a minha família, meus avós (*in memoriam*), minhas tias, tios, primos e padrinhos. Aos meus amigos, especialmente Lincoln, Marcos Wagner e Marcos Vianna, que se tornaram meus irmãos. Amo muito vocês. E também, aos meus colegas de turma, que se tornaram também grandes amigos.

Aos meus professores, em especial ao meu orientador, professor Wilton, muito obrigada pelos ensinamentos e pelo exemplo que é como profissional e pessoa. Ao meu coorientador, Marcos, que me ajudou e me incentivou, agradeço muito pela dedicação e parceria.

Agradeço também a todas as pessoas que me ajudaram na vida acadêmica, me deram oportunidades e fizeram com que eu acreditasse no meu potencial. Em especial, à Dra. Ana Grasielle, Dr. Elmar, Dr. Allancardi e Prof. Dr. Adriano.

RESUMO

Introdução: A análise cefalométrica se destaca pela sua importância na odontologia, especificamente na ortodontia, para diagnóstico e tratamento de má oclusão e discrepâncias esqueléticas. Recentemente, o traçado manual realizado por especialistas vem sendo substituído gradativamente por softwares que utilizam a inteligência artificial (IA). Dessa forma, torna-se relevante avaliar a precisão de sistemas automáticos para realizar esta tarefa. **Objetivo:** Avaliar, por meio de uma revisão sistemática (RS) com metanálise, a confiabilidade do uso de IA para a marcação cefalométrica em radiografias cefalométricas laterais. **Materiais e métodos:** Com o objetivo de responder ao questionamento “A marcação cefalométrica automática com IA é confiável?” foi realizada uma busca nas plataformas de pesquisas PubMed, Scopus, Embase, Web of Science, LILACS e Google Acadêmico, utilizando a estratégia de busca População, Intervenção, Comparação, Desfecho e Desenho de Estudo (PICOS). Os efeitos do tratamento foram definidos como diferença de média padronizada (SMD) e intervalos de confiança de 95% (IC) foram estabelecidos. Para avaliar o risco de viés, foi utilizado o questionário Joanna Briggs para estudos não randomizados. A ferramenta GRADE foi utilizada para avaliar a qualidade da evidência da revisão sistemática. **Resultados:** A estratégia de busca utilizada identificou um total de 1041 artigos. Destes, 32 foram selecionados em texto completo, e 14 foram incluídos na revisão sistemática após análise criteriosa. Destes 14, 5 estudos foram incluídos na metanálise para avaliação das medidas cefalométricas SNA, SNB, ANB e Wits. Nesta análise, obtivemos $Tau^2=0,04$, $Chi^2=31,70$, e valor de $p=0,01$, indicando a presença de heterogeneidade estatística, $I^2=50\%$, revelando uma heterogeneidade moderada. O SMD apresentou um valor total de $-0,05$, com intervalo de confiança de $-0,19$ a $0,08$, apontando que não há diferença estatística entre a medição automática feita por IA e a medição manual, porém, com um tamanho de efeito pequeno. **Conclusão:** A IA apresentou resultados semelhantes ao controle, na maior parte dos estudos acerca da precisão na identificação de pontos cefalométricos. Apesar da força de evidência da metanálise ser considerada muito baixa, os estudos incluídos apresentaram um baixo risco de viés.

Palavras-chave: Cefalometria, Inteligência Artificial, Software, Ortodontia.

ABSTRACT

Introduction: Cephalometric analysis stands out for its importance in dentistry, specifically in orthodontics, for the diagnosis and treatment of malocclusion and skeletal discrepancies. Recently, manual tracing performed by specialists has been gradually replaced by software that uses artificial intelligence (AI). Thus, it becomes relevant to assess the accuracy of automatic systems to perform this task. **Objective:** To evaluate, through a systematic review (SR) with meta-analysis, the reliability of the use of artificial intelligence for cephalometric marking. **Materials and methods:** In order to answer the question “Is automatic cephalometric marking with artificial intelligence reliable?” A search was carried out on the search platforms PubMed, Scopus, Embase, Web of Science, LILACS and Google Scholar, using the Population, Intervention, Comparison, Outcome and Study Design (PICOS) search strategy. Treatment effects were defined as standardized mean difference (SMD) and 95% confidence intervals (CI) were established. To assess the risk of bias, the Joanna Briggs questionnaire was used for non-randomized studies. The GRADE tool was used to assess the quality of evidence from the systematic review. **Results:** The search strategy used identified a total of 1041 articles. Of these, 32 were selected in full text, and 14 were included in the systematic review after careful analysis. Of these 14, 5 studies were included in the meta-analysis to assess macro cephalometric SNA, SNB, ANB and Wits. In this analysis, we obtained $\text{Tau}^2=0.04$, $\text{Chi}^2=31.70$, and $p \text{ value}=0.01$, indicating the presence of statistical heterogeneity, $I^2=50\%$, revealing moderate heterogeneity. The SMD presented a total value of -0.05, with a confidence interval of -0.19 to 0.09, indicating that there is no statistical difference between the automatic measurement performed by AI and the manual measurement, however, with a size of small effect. **Conclusion:** We obtained positive and promising results in most studies about the accuracy in identification of cephalometric landmarks. Although strength of evidence from meta-analysis was considered very low, included studies had a low risk of bias.

Keywords: Cephalometrics, Artificial Intelligence, Software, Orthodontics.

LISTA DE SIGLAS

ANB	Ângulo formado pela interseção das linhas NA e NB
GRADE	Grau de recomendação, Avaliação, Desenvolvimento e Análise
IA	Inteligência Artificial
JBÍ	Institute Joanna Briggs
PICOS	População, Intervenção, Comparação, Desfecho e Desenho do estudo
PROSPERO	International prospective register of systematic reviews
SD	Desvios padrão
SMD	Diferença de média padronizada
SNA	Ângulo obtido pela união da linha SN (sela ao ná시오) com a linha NA (ná시오 ao ponto A)
SNB	Ângulo obtido pela união da linha SN (sela ao ná시오) com a linha NB (ná시오 ao ponto B)
RevMan	Review Manager
Wits	Distância entre os pontos A e B perpendicular ao plano oclusal funcional

SUMÁRIO

1	INTRODUÇÃO	9
2	OBJETIVO	11
3	MATERIAIS E MÉTODOS.....	12
3.1.	Protocolo e registro.....	12
3.2.	Desenho do estudo e critérios de elegibilidade	12
3.3.	Fontes de pesquisa e informação	12
3.4.	Seleção de estudos	13
3.5.	Processo de coleta de dados	13
3.6.	Risco de Viés em estudos individuais.....	14
3.7.	Análise dos dados	14
3.8.	Qualidade da evidência	15
4	RESULTADOS.....	16
4.1.	Fontes de dados	16
4.2.	Características dos estudos.....	16
4.3.	Risco de viés dos estudos selecionados	17
4.4.	Síntese de dados.....	17
4.5.	Força da evidência	18
5	DISCUSSÃO	19
6	CONCLUSÃO	22
7	REFERÊNCIAS.....	23
8	TABELAS E FIGURAS.....	27
	Tabela 1. Banco de dados eletrônico e estratégia de busca	27
	Figura 1. Fluxograma do processo de seleção de artigos	28
	Tabela 2. Características dos estudos incluídos na Revisão Sistemática.....	29
	Figura 2. Gráfico de Floresta de manuscritos selecionados.....	33
	Tabela 3. Risco de viés avaliado pela ferramenta de avaliação crítica do Joanna Briggs Institute (JBI) para revisão sistemática de estudos observacionais	34
	Tabela 4. Perfil de evidência GRADE para estudos que avaliaram as medidas SNA, SNB, ANB e Wits.....	35

1 INTRODUÇÃO

Com o advento da cefalometria, várias análises cefalométricas foram criadas na tentativa de qualificar e quantificar os diversos componentes dentários e esqueléticos envolvidos nos diferentes tipos de más-oclusões, criando padrões e ideais cefalométricos. (THIENSEN, 2006) Apesar de ser o procedimento mais crucial e sensível na análise cefalométrica, a identificação de pontos de referência consome muito tempo, com alto potencial de erros e variabilidade. (YU et al., 2020) A precisão dos pontos de referência cefalométricos marcados pode afetar os resultados do desempenho clínico das análises e as decisões de tratamento resultantes. (LEE et al., 2020) O valor diagnóstico da análise depende da precisão e da reprodutibilidade da identificação do ponto de referência. (PARK et al., 2019)

Apesar do grande número de análises cefalométricas presentes na literatura ortodôntica, seus dados e parâmetros não são totalmente satisfatórios. Quando diferentes análises cefalométricas são utilizadas para avaliar um mesmo paciente, diferentes diagnósticos, planos de tratamento e resultados podem ser encontrados. (THIENSEN, 2006) Uma das causas mais significativas de erro de rastreamento é a incerteza na identificação do ponto de referência, que requer habilidades que dependem da experiência do examinador. (AKSAKALLI et al., 2016) Erros de inspeção manual devido à fadiga também podem aumentar a variabilidade intra-observador. (LEE et al., 2020)

Vários estudos propõem sistemas de identificação de fatores cefalométricos totalmente automatizados, baseados em uma técnica de aprendizado de máquina. (KIM et al., 2021) Em uma análise cefalométrica automatizada, uma radiografia cefalométrica digital é armazenada no computador e carregada pelo software. O software então localiza automaticamente os pontos de referência e realiza as medições para análise cefalométrica. (LEONARDI et al., 2008) Com as aplicações emergentes da IA em odontologia, pode-se esperar rastreamento e medição ainda mais rápidos, juntamente com maior precisão na anotação de pontos de referência. (MONTÚFAR et al., 2018)

Algoritmos baseados em IA estão incluídos na tecnologia do dia-a-dia e são amplamente utilizados, por exemplo, em mecanismos de busca na internet, filtragem de spam de e-mail ou assistentes online com reconhecimento de voz e até imagem em plataformas de mídia social. (KUNZ et al., 2020) O termo “inteligência artificial” (IA) foi cunhado na década de 1950 e refere-se à ideia de construir máquinas capazes de realizar tarefas que normalmente são realizadas por humanos. (SCHWENDICKE et al., 2020) A IA tem sido utilizada principalmente

na odontologia para tornar o processo de diagnóstico mais preciso e eficiente, o que é de suma importância para alcançar melhores resultados nos tratamentos prestados, juntamente com um atendimento de qualidade superior ao paciente. (KHANAGAR et al., 2021)

O aprendizado profundo demonstrou excelente desempenho em visão computacional, incluindo reconhecimento de objetos, faciais e de atividades, rastreamento e localização. Processamento de imagens e procedimentos de reconhecimento de padrões tornaram-se fatores-chave na segmentação e diagnóstico médico. (SHIN et al., 2021) Da mesma forma, muitos campos da medicina, incluindo câncer de mama e doenças pulmonares, mostraram usos bem-sucedidos de sistemas de diagnóstico direto baseados em rede neural convolucional com imagens de raios-x usadas isoladamente, sem informação adicional. (YU et al., 2020)

Os avanços tecnológicos levaram ao uso crescente de sistemas de análise cefalométrica digital, que apresentam várias vantagens: as doses de radiação são reduzidas, o armazenamento de dados é aprimorado e as imagens são facilmente manipuladas (AKSAKALLI et al., 2016). Os critérios mais importantes para o uso de método mecânico ou digital são que ele deve ser acurado, preciso e deve apresentar uma alta taxa de reprodutibilidade tanto no traçado quanto na análise. (FAROOQ et al., 2016)

Algoritmos publicados anteriormente podem ser divididos em duas categorias: floresta aleatória e rede neural convolucional. A árvore de decisão que compõe a floresta aleatória é uma estrutura lógica. A rede neural convolucional foi inspirada no mecanismo de reconhecimento visual natural de coisas vivas e é conhecido por ser mais adequado para processamento de imagem. (HWANG et al., 2021)

Desta forma, o objetivo deste trabalho é avaliar, por meio de uma revisão sistemática (RS) com metanálise, a confiabilidade do uso de IA para a marcação cefalométrica em radiografias cefalométricas laterais.

2 OBJETIVO

Avaliar, por meio de uma revisão sistemática com metanálise dos marcos cefalométricos SNA, SNB, ANB e Wits, a confiabilidade do uso de inteligência artificial para a marcação cefalométrica, em radiografias cefalométricas laterais.

3 MATERIAIS E MÉTODOS

3.1. Protocolo e registro

Esta revisão sistemática foi realizada de acordo com as diretrizes do protocolo PRISMA-P (www.prisma-statement.org/Protocols/) (Moher et al, 2015), com orientações do Manual Cochrane para revisões sistemáticas (<https://training.cochrane.org/handbook>) (Leeftang et al, 2013).

3.2. Desenho do estudo e critérios de elegibilidade

Com o objetivo de responder ao questionamento “A marcação cefalométrica automática com inteligência artificial é confiável?” foi realizada uma busca nas principais plataformas de pesquisas de artigos. Foi utilizada a estratégia de busca População, Intervenção, Comparação, Desfecho e Desenho de Estudo (PICOS).

Os estudos presentes nas buscas foram então avaliados e selecionados, de acordo com a temática e a metodologia aplicada. Alguns critérios de inclusão e exclusão foram utilizados, a fim de alcançar um conjunto de dados apropriado.

Os critérios de inclusão utilizados para esta revisão foram os seguintes: (1) estudos que avaliavam a precisão de marcação cefalométrica automática com uso de Inteligência Artificial, (2) estudos treinados e/ou testados em radiografias cefalométricas laterais 2D, (3) estudos que fizeram uma comparação entre o método automático proposto e o método manual, (4) estudos publicados a partir de 2015 até 2021.

Foram aplicados os seguintes critérios de exclusão: (1) estudos nos quais o objeto de interesse não era pertinente, (2) resumos ou índices, (3) cartas aos editores, (4) revisões de literatura, (5) informações pessoais ou comunicações curtas, (6) capítulos de livros, (7) patentes, (8) estudos com alto risco de viés ou de baixa qualidade metodológica, (9) teses e dissertações.

3.3. Fontes de pesquisa e informação

As bases de dados utilizadas foram as seguintes: PubMed, Scopus, Embase, Web of Science, LILACS e Google Acadêmico, utilizando-se as palavras-chave Cefalometry, Artificial

Intelligence, Deep learning, Accuracy. Os descritores foram selecionados com os Descritores em Ciências da Saúde (DeCS) e Medical Subject Headings (MeSH). Os operadores booleanos AND e OR foram utilizados nas estratégias para combinar palavras e auxiliar na melhor estratégia de busca a ser utilizada.

As buscas foram realizadas de forma sistêmica nas bases de dados escolhidas. Estratégias de busca diferentes foram utilizadas para cada base, a fim de obter o maior número de estudos possível (Tabela 1). A pesquisa abrangeu o período de janeiro de 2015 a junho de 2021.

Todas as referências obtidas a partir de PubMed, Scopus, Embase, Web of Science, LILACS e Google Acadêmico foram exportados para o software Microsoft Excel 2019, no qual registros duplicados foram removidos.

3.4. Seleção de estudos

Os dados foram coletados de forma independente por dois revisores: IMBSM e MALS. Em seguida, os títulos foram lidos com atenção para excluir artigos fora do escopo desta pesquisa. Os revisores não foram cegados por informações de autoria ou nome dos periódicos. Foram excluídos os estudos em que não foi possível abordar o assunto de interesse.

Inicialmente, os revisores selecionaram ou excluíram estudos com base no título e no resumo. Em seguida, foram verificados os critérios de inclusão e exclusão por meio de uma leitura superficial dos estudos. Por fim, numa terceira fase, os manuscritos que tinham sido selecionados foram lidos na íntegra, de forma independente pelos dois revisores, para avaliar se realmente continham os critérios de elegibilidade exigidos para participar da revisão sistemática. Estudos que incluíam análises em 3D através de tomografias computadorizadas, estudos que comparavam dois ou mais sistemas automatizados, ou estudos que não tinham como base o uso de um programa de Inteligência Artificial foram excluídos. Quando havia discrepância na opinião dos dois revisores, um terceiro revisor (WMT) contribuiu para um acordo final.

3.5. Processo de coleta de dados

Após a seleção dos estudos, o processo de coleta de dados foi realizado de forma sistemática. Os estudos foram revisados e uma tabela foi elaborada com suas principais

características, incluindo: autoria, ano de publicação, país de origem, número de especialistas que realizaram a medição manual, número de pontos e/ou marcações cefalométricas analisadas, número de radiografias para treinamento (dataset), número de radiografias para teste e tipo da estrutura de arquitetura utilizada.

3.6.Risco de Viés em estudos individuais

O risco de viés dos artigos selecionados foi investigado usando a ferramenta de Avaliação Crítica do Joanna Briggs Institute (JBI) para uso em estudos de revisão sistemática envolvendo a precisão do diagnóstico. Cada domínio relacionado ao risco potencial de viés foi avaliado de forma independente por dois revisores de acordo com as recomendações do PRISMA-P.

Foram usados os seguintes questionamentos através desta ferramenta: Q1) Está claro no estudo qual é a 'causa' e qual é o 'efeito' (ou seja, não há confusão sobre qual variável vem primeiro)?; Q2) Os participantes foram incluídos em comparações semelhantes?; Q3) Os participantes foram incluídos em quaisquer comparações que receberam tratamento/cuidado semelhante, além da exposição ou intervenção de interesse?; Q4) Havia um grupo de controle?; Q5) Houve múltiplas medições do resultado antes e após a intervenção/exposição?; Q6) O seguimento foi completo e, se não, foram as diferenças entre os grupos em termos de seu acompanhamento adequadamente descritos e analisados?; Q7) Os resultados dos participantes foram incluídos em quaisquer comparações medidas da mesma forma?; Q8) Os resultados foram medidos de forma confiável?; Q9) A análise estatística apropriada foi usada?. De acordo com a ferramenta, o risco de viés foi classificado como alto quando o estudo atingiu até 49% de pontuação “sim”, moderado quando atingiu 50 a 69% de pontuação “sim” e baixo quando atingiu mais de 70% de pontuação “sim”.

3.7.Análise dos dados

Uma metanálise de modelo de efeitos aleatórios foi realizada com os dados das medidas cefalométricas SNA, SNB, ANB e Wits, a fim de avaliar a precisão de detecção dos pontos entre os métodos manual e automático. O modelo de efeitos aleatórios utilizado foi o método Der Simonian-Laird, o qual determina a variância entre os estudos e, por conseguinte, descreve

a heterogeneidade entre os resultados dos estudos. A diferença média padronizada (SMD) foi utilizada como medidas de efeito e, para calcular o seu valor, médias e desvio padrão (SD) foram obtidas de cada grupo de estudo. O cálculo da estatística d de Cohen foi utilizado para determinar o tamanho do efeito, em que o valor de 0,2 foi considerado um efeito pequeno, o valor de 0,5 um efeito médio, e o valor de 0,8 um efeito grande.

Foram utilizados gráficos de floresta (*forest plot*) para representar os tamanhos de efeito e intervalos de confiança (CI) de 95%. Foi usado um $p < 0.05$ bicaudal para determinar a significância estatística. A heterogeneidade entre os estudos incluídos foi estimada usando o teste Q de Cochran e a estatística I^2 de Higgins & Thompson. As análises foram realizadas usando o software RevMan 5.3.

3.8. Qualidade da evidência

O software GRADEpro GDT (<http://gdt.guidelinedevelopment.org>) foi usado para resumir os resultados. A qualidade da evidência e a força das recomendações foram avaliadas usando a ferramenta Grau de Recomendação, Avaliação, Desenvolvimento e Avaliação (GRADE), com base no desenho do estudo, limitações metodológicas, inconsistência, evidência indireta, imprecisão e outras considerações, sendo classificado como alto, moderado, baixo e muito baixo. Estes níveis representam a confiança que possuímos na estimativa dos efeitos apresentados.

4 RESULTADOS

4.1.Fontes de dados

A estratégia de busca utilizada identificou um total de 1041 artigos. Destes, 32 artigos foram selecionados em texto completo, e 14 foram incluídos na revisão após análise criteriosa quanto à elegibilidade. O fluxograma do processo de seleção de artigos encontra-se na Figura 1.

4.2.Características dos estudos

As características dos estudos são apresentadas na Tabela 2. Os estudos incluídos variaram em número de marcos cefalométricos analisados, sendo o estudo de Hwang et al. (2020) o que apresentou a maior quantidade – 80 pontos cefalométricos – contra o estudo de Nishimoto et al. (2019), que apresentou a menor quantidade de pontos cefalométricos analisados – 10 pontos. Em relação ao número de examinadores, sete estudos apresentaram 2 examinadores para a realização da marcação cefalométrica manual, três estudos apresentaram 1 examinador, dois estudos apresentaram 3 examinadores, um estudo apresentou 12 examinadores e um estudo não relatou a quantidade de examinadores que realizaram a marcação manual.

Entre os estudos analisados, o que apresentou o maior número de dados para treinamento foi o estudo de Kim et al. (2021), com 3050 radiografias. O estudo com a menor quantidade de dados para treinamento foi de Vasamsetti et al. (2015), com 9 radiografias. Além disso, 4 estudos não apresentaram dados de treinamento, pois a comparação foi realizada com softwares de análise automática prontos para uso.

A fim de testar o método, o estudo de Lindner et al. (2016) utilizou 400 radiografias, a maior quantidade de dados entre os estudos incluídos, e os estudo de Anuwongnukroh et al. (2017) e Silva et al. (2021) utilizaram 30 radiografias, as menores quantidades observadas entre os estudos. Ademais, os tipos das estruturas de arquitetura da IA utilizadas nos estudos variaram, porém, a maioria deles utilizou as Redes Neurais Convolucionais, um algoritmo de aprendizado profundo que é utilizado no processamento e análise de imagens digitais.

4.3.Risco de viés dos estudos selecionados

A Tabela 3 apresenta a análise para risco de viés realizada através de uma lista de verificação para estudos experimentais não aleatórios, de acordo com as ferramentas de avaliação crítica para uso em revisões sistemáticas JBI. Todos os estudos incluídos apresentaram um risco incerto de viés em relação ao seu seguimento ser adequadamente descrito e analisado. Os estudos de Vasamsetti et al. (2015), Anuwongnukroh et al.(2017), Nishimoto et al. (2019), Kim et al. (2020) e Zeng et al. (2021) apresentaram problemas na medição confiável de seus resultados. Com isso, todos os estudos incluídos apresentaram um baixo risco de viés.

4.4.Síntese de dados

Para a metanálise, foram selecionados os estudos de Anuwongnukroh et al. (2017), Kunz et al. (2019), Nishimoto et al. (2019), Jeon and Lee (2021) e Silva et al. (2021). Estes estudos foram selecionados, pois apresentavam os fatores cefalométricos SNA, SNB, ANB e Wits, com valores de médias e desvio padrão. A Figura 2 apresenta o gráfico de floresta, utilizado para exibir graficamente os tamanhos de efeito e intervalos de confiança de 95%. Um $p < 0,05$ bicaudal foi usado para determinar a significância estatística. A heterogeneidade foi avaliada pelo teste Q de Cochran e quantificada pelo índice I^2 . Índices inferiores a 25% indicam baixa heterogeneidade entre os estudos, entre 25 e 75% heterogeneidade moderada e acima de 75% alta heterogeneidade.

Para a medida SNA, foram encontrados: $Tau^2=0,14$, $Chi^2=18,60$ e valor de $p=0.0009$, o que evidencia a presença de heterogeneidade estatística. Foi encontrado o valor de $I^2=78\%$, indicando uma heterogeneidade considerável. Para a medida SNB, foram encontrados: $Tau^2=0,00$, $Chi^2=1,10$ e valor de $p=0,89$, o que evidencia a homogeneidade dos estudos envolvidos. Foi encontrado o valor de $I^2=0\%$, indicando uma baixa heterogeneidade. Para a medida ANB, foram encontrados: $Tau^2=0,00$, $Chi^2=1,05$ e valor de $p=0,79$, o que evidencia a homogeneidade dos estudos envolvidos. Foi encontrado o valor de $I^2=0\%$, indicando uma baixa heterogeneidade. Para a medida Wits, foram encontrados: $Tau^2=0,00$, $Chi^2=0,35$ e valor de $p=0,84$, o que evidencia a homogeneidade dos estudos envolvidos. Foi encontrado o valor de $I^2=0\%$, indicando uma baixa heterogeneidade.

Em todas as metanálises realizadas, não há diferença estatística entre as duas intervenções (medição por IA e medição controle). Em uma análise total, obtivemos: $\text{Tau}^2=0,05$, $\text{Chi}^2=31,70$, e valor de $p=0,01$, indicando a presença de heterogeneidade estatística. Foi encontrado o valor de $I^2=50\%$, revelando uma heterogeneidade moderada. O SMD apresentou um valor total de $-0,05$, com intervalo de confiança de $-0,19$ a $0,08$, apontando que não há diferença estatística entre a medição automática feita por IA e a medição manual, porém, com um tamanho de efeito pequeno.

4.5. Força da evidência

A classificação baseada na ferramenta GRADE é apresentada na Tabela 4, caracterizada como uma tabela do tipo Perfil de Evidência. A força de evidência da confiabilidade da marcação cefalométrica realizada com IA foi classificada como alta, moderada, baixa ou muito baixa. Os estudos não randomizados iniciam como evidência de baixa qualidade, e, a partir disso, são verificados cinco fatores que podem diminuir ainda mais a evidência, são eles: risco de viés, inconsistência, evidência indireta, imprecisão e viés de publicação. Entretanto, a análise de três fatores pode aumentar o seu efeito: efeito dose-resposta, grande tamanho de efeito e confundidores residuais contrários ao benefício.

Com isso, a classificação encontrada para o uso da IA em marcação cefalométrica automática pelo GRADE foi considerada como muito baixa, indicando uma baixa confiança na estimativa desse efeito, resultado da análise feita por meio dos domínios que são orientados.

Para a medida SNA, a força de evidência foi considerada muito baixa, apresentando um SMD de $0,2$ e inconsistência muito grave. Para a medida SNB, a força de evidência foi classificada como baixa, com SMD de $0,04$. Para a medida ANB, a força de evidência foi baixa, com SMD de $0,08$. Por fim, para a medida Wits, a força de evidência foi considerada como baixa, apresentando um SMD de $0,13$.

5 DISCUSSÃO

A IA tem sido utilizada na odontologia principalmente como uma ferramenta para tornar o processo de diagnóstico mais preciso e eficiente, o que é de suma importância para alcançar melhores resultados nos tratamentos prestados, juntamente com um atendimento de qualidade superior ao paciente. (KHANAGAR et al., 2021) As radiografias cefalométricas laterais avaliam quantitativamente a relação esquelética entre a base do crânio, a maxila ou mandíbula, e a relação dento alveolar. (SCHWENDICKE et al., 2021) Por meio de radiografia cefalométrica padronizada, pontos anatômicos predefinidos são marcados para que várias análises ortodônticas e morfométricas faciais possam ser aplicadas para o diagnóstico e planejamento do tratamento. (LEE et al., 2020) Com a identificação imprecisa dos pontos cefalométricos pode levar a uma tomada de decisão incorreta para a terapia ortodôntica, uma identificação totalmente automatizada e confiável dos marcos cefalométricos é desejada, especialmente para fins de planejamento adequado. (ARIK et al., 2017) Com isso, esta revisão sistemática foi desenvolvida com o objetivo de avaliar a confiabilidade de abordagens de análise cefalométrica automática que utilizam IA.

Neste estudo, foi possível perceber o avanço que tem ocorrido na análise cefalométrica automática com o uso de IA, especialmente aquelas que utilizam as Redes Neurais Convolucionais (RNCs), um tipo de aprendizado profundo que é muito utilizado na detecção de objetos em imagens digitais. Identificamos alguns estudos que realizaram comparação entre um software de marcação automática já disponível comercialmente e a marcação manual: Anuwongnukroh et al. (2017), Anuwongnukroh et al. (2018), Silva, et al. (2021) e Jeon and Lee (2021). Os estudos restantes apresentaram uma nova abordagem de IA e testaram a sua precisão. Vale destacar que o estudo de Lindner et al. (2016) modificou e aprimorou um sistema de identificação de pontos de referência em ossos do fêmur para realizar a identificação de pontos cefalométricos em cefalogramas laterais.

A maior parte dos estudos apresentaram resultados favoráveis em relação à precisão de análise. Isso corrobora com os achados de Schwendicke et al. (2021), em que se mostra uma precisão considerável para estudos que utilizam aprendizagem profunda na detecção de pontos de referência em imagens cefalométricas. Na presente revisão, Anuwongnukroh et al. (2017) e Anuwongnukroh et al. (2018) avaliaram o seu sistema automático como imprecisos; Nishimoto et al. (2018) destacaram que, apesar dos bons resultados, o sistema ainda estava sendo desenvolvido e havia necessidade de avaliação clínica; Jeon e Lee (2021), Silva et al. (2021) e

Kim et al. (2021) consideraram seus sistemas promissores, porém, enfatizaram a importância da supervisão humana; Vasamsetti et al. (2015), Linder et al. (2016), Arik et al. (2017), Kunz et al. (2019), Hwang et al. (2020), Kim et al. (2020), Lee et al. (2020) e Zeng et al. (2021) consideraram seus sistemas como precisos e confiáveis, compatíveis com o ambiente clínico atual.

De acordo com a avaliação do risco de viés através da ferramenta JBI, obtivemos um baixo risco de viés entre os estudos, como pode ser visto na Tabela 3. As respostas ao questionário foram positivas em aproximadamente 85% dos casos. Em relação ao questionamento acerca do seguimento, não ficou claro se havia esse acompanhamento, já que eram estudos associados à precisão de medidas. Além disso, alguns estudos não utilizaram ou não descreveram informações referentes à confiabilidade da medição de resultados, a exemplo de calibração entre examinadores e confiabilidade inter e intra-examinadores.

Para a metanálise, utilizamos as medidas SNA, SNB, ANB e Wits, que estavam presentes em pelo menos 3 estudos incluídos. Na Figura 2, podemos observar o gráfico de floresta de cada medida avaliada. Para a medida SNA, temos 5 estudos incluídos, com 4 deles não atravessando a linha do eixo zero, apenas o trabalho de Anuwongnukroh et al. (2017) apresentou diferença estatística, que pode ser justificado pelo fato de terem utilizados cefalogramas e não radiografias cefalométricas laterais, os algoritmos de aprendizado de máquina são melhores treinados, quando utilizados com imagens com maior quantidade de detalhes. Apesar disso, em uma abordagem geral, o diamante não atravessou o eixo zero, o que nos revela que não há diferença estatística entre as duas intervenções (IA e controle) para esta medida.

Para SNB, temos 5 estudos inseridos no gráfico, onde todos não cruzam a linha do eixo zero, incluindo o diamante, indicando que não há diferença estatística entre as duas intervenções. Para ANB, 4 estudos foram inseridos no gráfico e todos não atravessaram a linha do eixo 0, assim como o diamante, evidenciando que nesta medida também não há diferença estatística as intervenções. Para Wits, temos 3 estudos incluídos no gráfico que não atravessaram a linha do eixo zero, assim como o diamante, o que indica que não há diferença estatística entre as intervenções para esta medida. Por fim, o diamante que representa a junção das metanálises indica que não há diferença estatística entre as intervenções para os marcos analisados.

Em relação ao resultado da ferramenta GRADE, nosso estudo obteve uma força de evidência considerada muito baixa. Isso pode ser explicado pelo fato de que os estudos

incluídos são do tipo observacional, assim como pela presença de inconsistência, que foi considerada como grave. A inconsistência indica uma heterogeneidade, que dificulta a interpretação dos resultados e indica que houve diferenças nas estimativas de efeito de intervenções entre os estudos. O marco cefalométrico SNA obteve o maior valor de I^2 , de 78%, revelando alta heterogeneidade, resultado oposto aos outros marcos (SNB, ANB e Wits) que apresentaram $I^2=0\%$, mas impactando o resultado final. Além disso, é válido destacar que esta avaliação foi feita com base apenas nos 5 artigos presentes na metanálise.

Em 2008, Leonardi et. al publicaram uma revisão sistemática com a proposta de avaliar a porcentagem de sucesso na detecção de pontos cefalométricos por marcação automática, e sua conclusão foi a de que os sistemas descritos não eram suficientemente precisos para serem utilizados clinicamente. Em 2021, após muitos avanços na área da IA, aproximadamente treze anos se passaram e Schwendicke et al. (2021) realizaram uma revisão sistemática com o objetivo de avaliar a precisão de sistemas que utilizaram aprendizagem profunda na marcação cefalométrica automática em radiografias 2-D e 3-D. Sua conclusão foi a de que os estudos apresentam uma precisão relativamente alta para a detecção de pontos cefalométricos, o que corrobora com a avaliação que obtivemos neste estudo. Contudo, nosso estudo foi pioneiro, pois até o presente momento foi o primeiro a comparar as medidas cefalométricas SNA, SNB, ANB e Wits. Entendemos que os estudos favorecem a marcação do ponto cefalométrico, contudo no intuito de conferir maior pragmatismo ao estudo a comparação por meio de medidas cefalométricas parece ser um caminho promissor.

Alguns artigos admitiram que uma forma de definir o sucesso é por meio do número ou percentual de pontos totais detectados dentro do limite de 2 mm. (VASAMSETTI et al., 2015; ARIK et al., 2017; HWANG et al., 2020; LEE et al., 2020). Anteriormente, a margem de erro era alta, com erros máximos variando entre 4,4 e 7,3 mm, portanto, a aplicação clínica não era possível. (KIM et al., 2021) Ao longo da última década, vários campos clínicos relataram um aumento na eficiência clínica de acordo com a aplicação da inteligência artificial. (LEE et al., 2020) Recentemente, métodos automatizados de análise cefalométrica usando aprendizado de máquina e aprendizado profundo foram introduzidos e ganharam atenção por sua precisão aprimorada. (KIM et al., 2021)

Em relação às limitações desta revisão, selecionamos apenas estudos que utilizaram radiografias 2D, não havendo uma comparação com imagens 3D; incluímos estudos a partir de 2015, um ano a partir do qual consideramos que a IA estava evoluindo consideravelmente, porém, não avaliamos como foi o progresso desses sistemas através de estudos anteriores;

alguns estudos foram excluídos pois as formas de avaliação de precisão não eram semelhantes em todos os estudos envolvidos; realizamos uma metanálise com uma quantidade limitada de medições cefalométricas; além disso, a força de evidência foi considerada baixa, o que pode indicar a necessidade de ensaios clínicos randomizados para elevar esta classificação.

Para futuras pesquisas, há a necessidade de englobar imagens 2-D e 3-D, realização de metanálises que atinjam um maior número de medições cefalométricas e inclusão de novos ensaios clínicos randomizados para dar maior força à evidência.

6 CONCLUSÃO

Neste trabalho, obtivemos resultados positivos e promissores na maior parte dos estudos acerca da precisão na identificação de pontos cefalométricos. A comparação das medidas SNA, SNB, ANB e Wits estudadas apresentaram-se como semelhantes entre a IA e o controle. Apesar da força de evidência da metanálise ser considerada muito baixa, os estudos incluídos apresentaram um baixo risco de viés. Novas pesquisas devem ser realizadas a fim de elevar a qualidade da evidência para atestar a confiabilidade dos sistemas automáticos que utilizam a IA em sua estrutura.

7 REFERÊNCIAS

1. Aksakallı S, Yılandı H, Görükmez E, Ramoğlu Sİ. Reliability Assessment of Orthodontic Apps for Cephalometrics. *Turk J Orthod*. 2016 Dec;29(4):98-102. doi: <https://doi.org/10.5152/TurkJOrthod.2016.1618>.
2. Anuwongnukroh N, et al. Accuracy of Automatic Cephalometric Software on Landmark Identification. *IOP Conf. Series: Materials Science and Engineering*, 265 (2017) 012028. doi: <https://doi.org/10.1088/1757-899X/265/1/012028>.
3. Anuwongnukroh N, et al. Assessment of the Reliability of Automatic Cephalometric Analysis Software. *International Journal of Mechanical Engineering and Robotics Research*, vol. 7, no. 1, January 2018. doi: <https://doi.org/10.18178/ijmerr.7.1.61-65>.
4. Braga, A. V. et al. Machine learning: The Use of Artificial Intelligence in Medicine. *Braz. J. of Develop.*, Curitiba, v. 5, n. 9, p. 16407-16413 sep. 2019. doi: <https://doi.org/10.34117/bjdv5n9-190>.
5. Brasil. Ministério da Saúde. Secretaria de Ciência, Tecnologia, Inovação e Insumos Estratégicos em Saúde. Departamento de Gestão e Incorporação de Tecnologias em Saúde. Diretrizes metodológicas: elaboração de revisão sistemática e meta-análise de ensaios clínicos randomizados [recurso eletrônico]. Brasília: Ministério da Saúde, 2021. 93 p.
6. Farooq, M.U., Khan, M.A., Imran, S., Sameera, A., Reshi, A.Q., Ahmed, S.A., Kumar, S., Rahman, M.A., (2016). Assessing the Reliability of Digitalized Cephalometric Analysis in Comparison with Manual Cephalometric Analysis. *J Clin of Diagn Res*. 10(10), ZC20-ZC23. doi: <https://www.doi.org/10.7860/JCDR/2016/17735/8636>.
7. Hye-Won Hwang, Jun-Ho Moon, Min-Gyu Kim, Richard E. Donatelli, Shin-Jae Lee; Evaluation of automated cephalometric analysis based on the latest deep learning method. *Angle Orthod*, 2021; 91 (3): 329–335. doi: <https://doi.org/10.2319/021220-100.1>.
8. Hye-Won Hwang et al. Automated Identification of Cephalometric Landmarks: Part 2- Might It Be Better Than human?. *Angle Orthod*, 14 January 2020; 90 (1): 69–76. doi: <https://doi.org/10.2319/022019-129.1>.
9. H. Kim, E. Shim and J. Park et al. Web-based fully automated cephalometric analysis by deep learning. *Computer Methods and Programs in Biomedicine* 194 (2020) 105513. doi: <https://doi.org/10.1016/j.cmpb.2020.105513>.

10. Jeon, S., Lee, K.C. Comparison of cephalometric measurements between conventional and automatic cephalometric analysis using convolutional neural network. *Prog Orthod.* 22, 14 (2021). doi: <https://doi.org/10.1186/s40510-021-00358-4>.
11. Jaerong Kim, et al. Accuracy of automated identification of lateral cephalometric landmarks using cascade convolutional neural networks on lateral cephalograms from nationwide multi-centres. *Orthod Craniofac Res.* 2021;00:1–9. doi: <https://doi.org/10.1111/ocr.12493>.
12. Khanagar, S. B. et al. Developments, application, and performance of artificial intelligence in dentistry - A systematic review. *Journal of Dental Sciences* (2021) 16, 508-522. doi: <https://doi.org/10.1016/j.jds.2020.06.019>.
13. Kim, IH., Kim, YG., Kim, S. et al. Comparing intra-observer variation and external variations of a fully automated cephalometric analysis with a cascade convolutional neural net. *Sci Rep* 11, 7925 (2021). doi: <https://doi.org/10.1038/s41598-021-87261-4>.
14. Kim M-J, Liu Y, Oh SH, Ahn H-W, Kim S-H, Nelson G. Automatic Cephalometric Landmark Identification System Based on the Multi-Stage Convolutional Neural Networks with CBCT Combination Images. *Sensors.* 2021; 21(2):505. doi: <https://doi.org/10.3390/s21020505>.
15. Kunz, F., Stellzig-Eisenhauer, A., Zeman, F. et al. Artificial intelligence in orthodontics: Evaluation of a fully automated cephalometric analysis using a customized convolutional neural network. *J Orofac Orthop*, 81, 52–68 (2020). doi: <https://doi.org/10.1007/s00056-019-00203-8>.
16. Lee, JH., Yu, HJ., Kim, Mj. et al. Automated cephalometric landmark detection with confidence regions using Bayesian convolutional neural networks. *BMC Oral Health* 20, 270 (2020). doi: <https://doi.org/10.1186/s12903-020-01256-7>.
17. Leonardi, R., Giordano, D., Maiorana, F., Spampinato, C. Automatic Cephalometric Analysis: A Systematic Review. *Angle Orthodontist*, vol 78, no 1, 2008. doi: <https://doi.org/10.2319/120506-491.1>.
18. Leonardi, R., Giordano, D., Maiorana, F. An Evaluation of Cellular Neural Networks for the Automatic Identification of Cephalometric Landmarks on Digital Images, *BioMed Research International*, vol. 2009, Article ID 717102, 12 pages, 2009. doi: <https://doi.org/10.1155/2009/717102>.

19. Lindner, C., Wang, CW., Huang, CT. et al. Fully Automatic System for Accurate Localisation and Analysis of Cephalometric Landmarks in Lateral Cephalograms. *Sci Rep* 6, 33581 (2016). doi: <https://doi.org/10.1038/srep33581>.
20. Meriç P, Naoumova J. Web-based Fully Automated Cephalometric Analysis: Comparisons between App-aided, Computerized, and Manual Tracings. *Turk J Orthod* 2020; 33(3): 142-9. doi: <https://doi.org/10.5152/TurkJOrthod.2020.20062>.
21. Montúfar J, Romero M, Scougall-Vilchis RJ. Automatic 3-dimensional cephalometric landmarking based on active shape models in related projections. *Am J Orthod Dentofacial Orthop* 2018; 153: 449–58. doi: <https://doi.org/10.1016/j.ajodo.2017.06.028>.
22. M. Zeng, Z. Yan, S. Liu et al. Cascaded convolutional networks for automatic cephalometric landmark detection. *Medical Image Analysis*. 2021 Feb;68:101904. doi: <https://doi.org/10.1016/j.media.2020.101904>.
23. Neto, C. D. N. et. al. Artificial intelligence and new health technologies: Challenges and prospects. *Braz. J. of Develop.*, Curitiba, v. 6, n. 2, p. 9431-9445, feb. 2020. doi: <https://doi.org/10.34117/bjdv6n2-306>.
24. Park, J-H. et al. Automated identification of cephalometric landmarks: Part 1—Comparisons between the latest deep-learning methods YOLOV3 and SSD. *Angle Orthodontist*, vol 89, no 6, 2019. doi: <https://doi.org/10.2319/022019-127.1>.
25. Schwendicke F, Samek W, Krois J. Artificial Intelligence in Dentistry: Chances and Challenges. *Journal of Dental Research*. 2020;99(7):769-774. doi: <https://doi.org/10.1177/0022034520915714>.
26. Schwendicke, F., Chaurasia, A., Arsiwala, L. et al. Deep learning for cephalometric landmark detection: systematic review and meta-analysis. *Clin Oral Invest* 25, 4299–4309 (2021). <https://doi.org/10.1007/s00784-021-03990-w>
27. Sercan Ö. Arık, Bulat Ibragimov, Lei Xing. Fully automated quantitative cephalometry using convolutional neural networks. *J. Med. Imag.* 4(1), 014501 (2017), doi: <https://doi.org/10.1117/1.JMI.4.1.014501>.
28. Sertaç Aksakallı, Hilal Yılcı, Erhan Görükmez , Sabri İlhan Ramoğlu. Reliability Assessment of Orthodontic Apps for Cephalometrics, *Turkish J Orthod* 2016; 29: 98-102. doi: <https://doi.org/10.5152/TurkJOrthod.2016.1618>.

29. Shin, W., Yeom, HG., Lee, G.H. et al. Deep learning based prediction of necessity for orthognathic surgery of skeletal malocclusion using cephalogram in Korean individuals. *BMC Oral Health*, 21, 130 (2021). doi: <https://doi.org/10.1186/s12903-021-01513-3>.
30. Silva, T. et al. Artificial intelligence-based cephalometric landmark annotation and measurements according to Arnett's analysis: can we trust a bot to do that? *Dentomaxillofacial Radiology* (2021) 50, 20200548. doi: <https://doi.org/10.1259/dmfr.20200548>.
31. Soh Nishimoto et al. Personal computer-based cephalometric landmark detection with deep learning, using cephalograms on the internet, *Journal of Craniofacial Surgery*: January/February 2019, Volume 30, Issue 1, p 91-95 doi: <https://doi.org/10.1097/SCS.00000000000004901>.
32. Thiensen, G. The role of facial analysis in contemporary orthodontics. *Rev Assoc Paul Cir Dent*;60(5):380-, 2006. ID: lil-543819.
33. Vasamsetti, S. et al. Automatic landmark identification in lateral cephalometric images using optimized template matching. *Journal of Medical Imaging and Health Informatics*, Volume 5, Number 3, June 2015, pp. 458-470(13). doi: <https://doi.org/10.1166/jmihi.2015.1426>.
34. Yu, H. J. et al. Automated Skeletal Classification with Lateral Cephalometry Based on Artificial Intelligence. *Journal of Dental Research*, vol. 99(3) 249–256, 2020. doi: <https://doi.org/10.1177/0022034520901715>.

8 TABELAS E FIGURAS

Tabela 1. Banco de dados eletrônico e estratégia de busca

<i>Banco de dados</i>	<i>Estratégia de busca (junho de 2021)</i>
PubMed https://www.ncbi.nlm.nih.gov/pubmed/	(cephalometry OR cephalometric tracing OR cephalometry 2D) AND (artificial intelligence OR deep learning OR software)
Web of Science https://www.webofscience.com	((ALL=(cephalometric tracing)) AND ALL=(software)) AND ALL=(accuracy)
Embase http://www.embase.com	(cephalometry) AND (deep learning)
Scopus http://www.scopus.com/	Cephalometry AND "Artificial intelligence" AND accuracy
Lilacs https://lilacs.bvsalud.org/	“Cefalometria” and “inteligência artificial”
Google Scholar https://scholar.google.com.br/	cephalometry OR cephalometry 2D OR cephalometric tracing OR cephalometric landmark OR cephalometric marking AND artificial intelligence OR deep learning OR software OR artificial neural network AND data accuracy OR accuracy

Figura 1. Fluxograma do processo de seleção de artigos

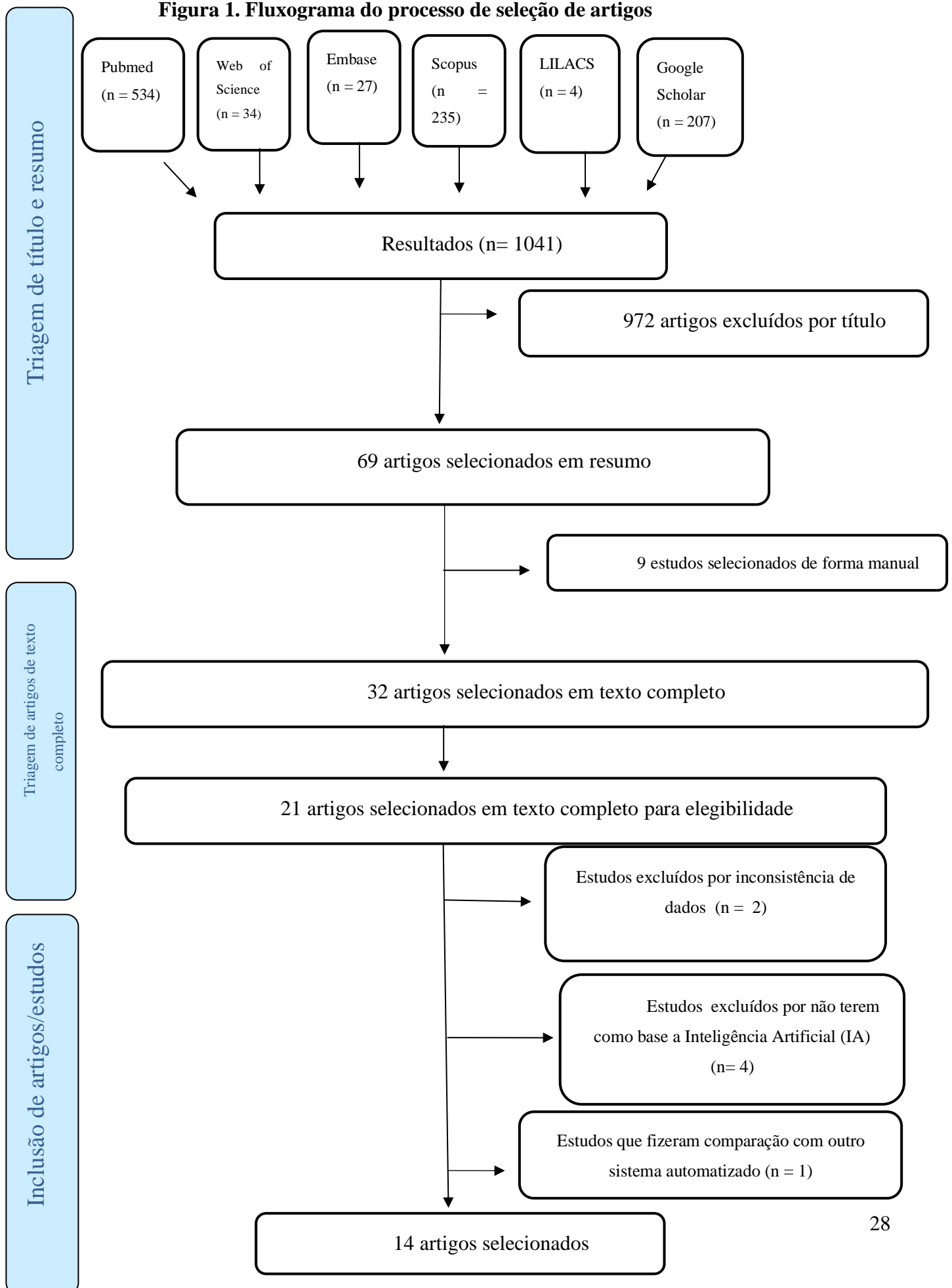


Tabela 2. Características dos estudos incluídos na Revisão Sistemática

Autores	Ano	País	Número de pontos cefalométricos analisados	Número de examinadores	Número de radiografias para treinamento (dataset)	Número de radiografias para teste	Estrutura de arquitetura	Conclusão
Srikanth Vasamsetti et al.	2015	Índia	24	3 especialistas ortodônticos.	9	37	Algoritmo de correspondência de modelos otimizado (OTMA) usando o ambiente de programação MATLAB.	O método proposto foi capaz de alcançar acurácia e precisão significativas na detecção de pontos de referência, juntamente com a relevância clínica de identificar os pontos de referência declarados e, assim, avançar um passo na realização de análises cefalométricas computadorizadas de maneira automatizada.
Claudia Lindner et al.	2016	Taiwan	19	2 ortodontistas clínicos (com seis anos e 15 anos de experiência, respectivamente).	150	400	Votação de regressão Random Forest usada tanto para detectar a posição, escala e orientação do crânio e, em seguida, estrutura Constrained Local Model (RFRV-CLM), para localizar os pontos de referência individuais.	O sistema FALA mostra uma grande promessa para aplicação em soluções de software ortodôntico para conduzir análises cefalométricas de forma totalmente automática.
Arik et al.	2017	EUA	19	2 examinadores.	150	250	Redes neurais convolucionais profundas.	No geral, nossa estrutura demonstra alta precisão de detecção de pontos anatômicos e alta precisão de classificação de tipo anatômico. Espera-se que os resultados melhorem ainda mais com o aumento na quantidade de conjunto de dados de treinamento.
Anuwongnukroh N. et al.	2017	Tailândia	13	2 ortodontistas experientes.	-	30	Software de imagem dentária (Carestream Dental, versão 6.14), programa de análise cefalométrica totalmente	Os resultados sugeriram que a marcação computadorizada desse software não era precisa o suficiente

Niwat Anuwongnukroh et al.	2018	Tailândia	17	2 examinadores.	-	108	automático disponível na Mahidol University. Software de imagem dentária (Carestream Dental, versão 6.14)	para permitir seu uso para fins clínicos. O modo totalmente automático do software de análise cefalométrica não é tão confiável quanto a análise manual. Ele deve ser usado apenas para apoiar um diagnóstico e não como uma ferramenta de diagnóstico.
Felix Kunz et al.	2019	Alemanha	18 (foram usados para representar 12 parâmetros)	12 examinadores (6 ortodontistas, 6 dentistas na segunda metade de sua pós-graduação em ortodontia).	1792	50 (porém, 3 apresentaram artefatos graves e foram excluídas da análise estatística)	Algoritmo de aprendizado profundo da CNN (rede neural convolucional) de código aberto personalizado (Keras e Google Tensorflow).	Devido ao alto nível de qualidade e à grande quantidade de dados de treinamento, conseguimos gerar um algoritmo de IA capaz de analisar novas radiografias cefalométricas com precisão comparável a examinadores humanos experientes, o que é considerado o padrão ouro atual.
Soh Nishimoto et al.	2019	Japão	10	-	153	66	Python 3.5 (Python Software Foundation, Beaverton, OR): uma linguagem de programação foi usada no Anaconda como um sistema de instalação e Spyder 3.5 como um ambiente de desenvolvimento integrado. Keras: a biblioteca de aprendizado profundo, escrita em Python, foi executada no TensorFlow (Google, Mountain View, CA).	Os ângulos e comprimentos na análise cefalométrica, previstos pela rede neural, não foram significativamente diferentes daqueles calculados pelos valores das coordenadas e plotados manualmente.
Hye-Won Hwang et al.	2020	Coreia do Sul	80	1 examinador com 28 anos de experiência clínica ortodôntica.	1028	283	Aprendizado de máquina – Deep learning (aprendizado profundo). Algoritmo You-Only-Look-Once versão 3 (YOLOv3)	O IA proposto no presente estudo pode ser compatível com o ambiente clínico atual e manter sua validade sob a supervisão constante de especialistas em ortodontia.
Kim, H. et al.	2020	Korea	23	2 ortodontistas.	2075	100	Deep learning. Todos os servidores foram implementados com Flask em Python. As páginas web do cliente foram implementadas	Os 23 pontos cefalométricos foram identificados automaticamente com alta precisão imediatamente, e o algoritmo proposto obteve resultados muito semelhantes à verdade básica

							com React em JavaScript, e o banco de dados foi construído com MongoDB. Rede de ampulheta empilhada (SHG). Redes Neurais Convolucionais Bayesianas (BCNN).	para a classificação dos tipos anatômicos.
Jeong-Hoon Lee et al.	2020	Korea	19	2 especialistas.	150	250		Nossa estrutura pode servir como uma ferramenta de diagnóstico auxiliada por computador que melhora a precisão e a confiabilidade das decisões dos especialistas.
Silva, T. P. et al.	2021	Brasil	66 + 10 medidas lineares e angulares.	1 radiologista dentário com mais de 20 anos de experiência em Cefalometria computadorizada.	-	30	Software baseado em aprendizado de máquina (ML)-CNN.	Nossos resultados sugerem que o software CEFBOT baseado em IA é, em sua versão atual, uma ferramenta promissora para identificação e marcação de pontos cefalométricos de acordo com a análise de Arnett - desde que seja usado sob a supervisão de um radiologista.
Sangmin Jeon and Kyungmin Clara Lee	2021	Coréia	42	1 examinador experiente.	-	35	Rede Neural Convolucional (CNN). Programa Ceph-X.	Com a limitação deste estudo, os resultados indicam que análises cefalométricas automáticas baseadas em rede neural convolucional podem oferecer desempenho diagnóstico clinicamente aceitável.
Jaerong Kim et al.	2021	Coréia	20	2 ortodontistas com 10 e 30 anos de experiência clínica, e 1 residente ortodôntico do segundo ano.	3050 1392+440	100	Rede Neural Convolucional em cascata.	Embora não possamos concluir que este modelo de detecção automatizada de pontos de referência pode substituir a tarefa humana de análise cefalométrica lateral, ele pode ajudar na triagem preliminar para o diagnóstico do paciente e avaliação do meio do tratamento, independente do tipo das máquinas de radiografia testadas.
Minmin Zeng et al.	2021	China	19	2 examinadores experientes.	400	150	Rede neural convolucional em cascata.	Os resultados da previsão neste conjunto de dados mostram que nossa abordagem pode ser considerada como uma abordagem prática baseada

Figura 2. Gráfico de Floresta de manuscritos selecionados

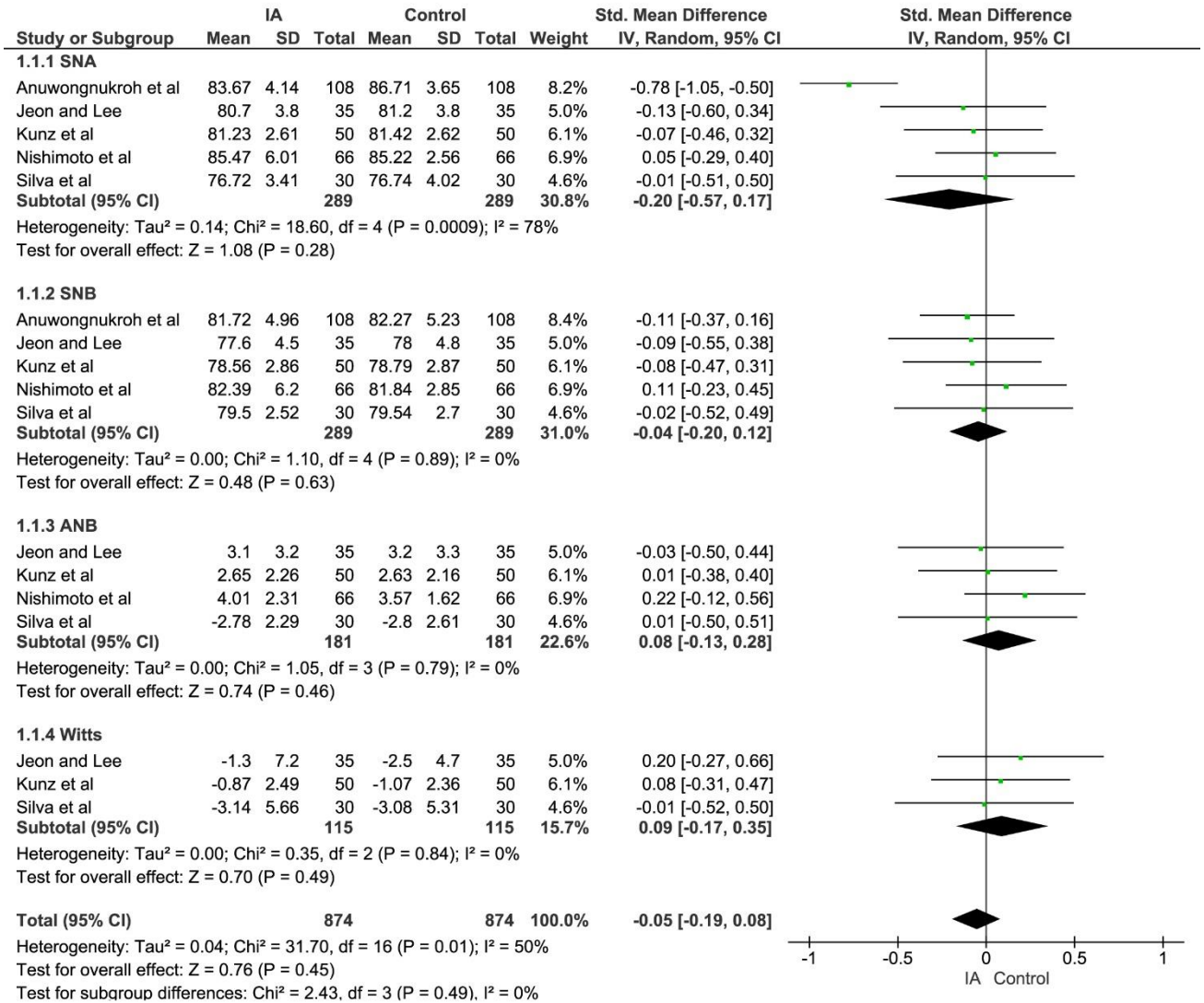


Tabela 3. Risco de viés avaliado pela ferramenta de avaliação crítica do Joanna Briggs Institute (JBI) para revisão sistemática de estudos observacionais

Autores	Ano	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	% Sim	Risco de Viés
Vasamsetti et al.	2015	✓	✓	✓	✓	✓	U	✓	X	✓	77,8%	BAIXO
Lindner et al.	2016	✓	✓	✓	✓	✓	U	✓	✓	✓	88,9%	BAIXO
Arik et al.	2017	✓	✓	✓	✓	✓	U	✓	✓	✓	88,9%	BAIXO
Anuwongnukroh et al.	2017	✓	✓	✓	✓	✓	U	✓	X	✓	77,8%	BAIXO
Nilwarat et al.	2018	✓	✓	✓	✓	✓	U	✓	✓	✓	88,9%	BAIXO
Kunz et al.	2019	✓	✓	✓	✓	✓	U	✓	✓	✓	88,9%	BAIXO
Nishimoto et al.	2019	✓	✓	✓	✓	✓	U	✓	X	✓	77,8%	BAIXO
Hwang et al.	2020	✓	✓	✓	✓	✓	U	✓	✓	✓	88,9%	BAIXO
Kim et al.	2020	✓	✓	✓	✓	✓	U	✓	X	✓	77,8%	BAIXO
Lee, et al.	2020	✓	✓	✓	✓	✓	U	✓	✓	✓	88,9%	BAIXO
Silva et al.	2021	✓	✓	✓	✓	✓	U	✓	✓	✓	88,9%	BAIXO
Jeon and Lee	2021	✓	✓	✓	✓	✓	U	✓	✓	✓	88,9%	BAIXO
Kim et al.	2021	✓	✓	✓	✓	✓	U	✓	✓	✓	88,9%	BAIXO
Zeng et al.	2021	✓	✓	✓	✓	✓	U	✓	X	✓	77,8%	BAIXO
											% Média	84,9% BAIXO

Tabela 4. Perfil de evidência GRADE para estudos que avaliaram as medidas SNA, SNB, ANB e Wits

Certainty assessment							N: of patients		Effect		Certainty	Importance
N: of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Cephalometry IA	placebo	Relative (95% CI)	Absolute (95% CI)		
Cephalometry IA												
5	observational studies	not serious	serious	not serious	not serious	none	814	874	-	SMD 0.05 lower (0.21 lower to 0.1 higher)	⊕○○○ Very low	
Cephalometry IA - SNA												
5	observational studies	not serious	very serious	not serious	not serious	none	289	289	-	SMD 0.2 lower (0.57 lower to 0.17 higher)	⊕○○○ Very low	
Cephalometry IA - SNB												
5	observational studies	not serious	not serious	not serious	not serious	none	259	289	-	SMD 0.04 lower (0.22 lower to 0.13 higher)	⊕⊕○○ Low	
Cephalometry IA - ANB												
4	observational studies	not serious	not serious	not serious	not serious	none	181	181	-	SMD 0.08 higher (0.13 lower to 0.28 higher)	⊕⊕○○ Low	
Cephalometry IA - Witts												
3	observational studies	not serious	not serious	not serious	not serious	none	85	115	-	SMD 0.13 higher (0.17 lower to 0.43 higher)	⊕⊕○○ Low	

CI: confidence interval; **SMD**: standardised mean difference