

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE ODONTOLOGIA

JERUSA JOBIM JARDIM

INTELIGÊNCIA ARTIFICIAL APLICADA AOS EXAMES
DE IMAGEM ODONTOLÓGICOS

Porto Alegre

2021

JERUSA JOBIM JARDIM

INTELIGÊNCIA ARTIFICIAL APLICADA AOS EXAMES
DE IMAGEM ODONTOLÓGICOS

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Radiologia Odontológica e Imaginologia da Faculdade de Odontologia da Universidade Federal do Rio Grande do Sul como requisito parcial para obtenção do título de Especialista em Radiologia Odontológica.

Orientadora: Prof^a Dr^a Nádia Assein Arús – Especialista em Radiologia Oral (UFRJ), Mestre e Doutora em Clínica Odontológica/Radiologia (UFRGS)

Porto Alegre

2021

CIP - Catalogação na Publicação

JARDIM, JERUSA JOBIM
INTELIGÊNCIA ARTIFICIAL APLICADA AOS EXAMES DE
IMAGEM ODONTOLÓGICOS / JERUSA JOBIM JARDIM. -- 2021.
73 f.
Orientadora: NÁDIA ASSEIN ARÚS.

Trabalho de conclusão de curso (Especialização) --
Universidade Federal do Rio Grande do Sul, Faculdade
de Odontologia, ESPECIALIZAÇÃO EM RADIOLOGIA
ODONTOLÓGICA E IMAGINOLOGIA , Porto Alegre, BR-RS,
2021.

1. INTELIGÊNCIA ARTIFICIAL. 2. DIAGNÓSTICO POR
IMAGEM. 3. ODONTOLOGIA. I. ARÚS, NÁDIA ASSEIN, orient.
II. Título.

JERUSA JOBIM JARDIM

INTELIGÊNCIA ARTIFICIAL APLICADA AOS EXAMES
DE IMAGEM ODONTOLÓGICOS

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Radiologia Odontológica e Imaginologia da Faculdade de Odontologia da Universidade Federal do Rio Grande do Sul como requisito parcial para obtenção do título de Especialista em Radiologia Odontológica.

Aprovado em: ____ de _____ de _____.

Profª Drª Nádia Assein Arús
Especialista em Radiologia Oral (UFRJ), Mestre e Doutora em Clínica
Odontológica/Radiologia (UFRGS)

Profª Drª Gabriela Salatino Liedke
Mestre e Doutora em Clínica Odontológica/Radiologia (UFRGS)

Profª Drª Niége Michelle Lazzari Onofre
Especialista em Ortodontia (PUC), Mestre em Ortodontia (PUC) e Doutora
em Clínica Odontológica/Radiologia (UFRGS)

Porto Alegre

2021

RESUMO

A presente revisão da literatura teve por objetivo investigar a literatura relacionada à aplicação da Inteligência Artificial (IA) na análise de exames de imagem nas diversas especialidades odontológicas, seu desempenho nesta tarefa, possibilidades de aplicação na prática clínica, bem como suas vantagens e limitações. A busca bibliográfica realizada em uma fonte de dados resultou em inclusão de 58 artigos originais de texto completo, trabalhos experimentais e revisões sistemáticas da literatura, publicados em inglês, período de publicação entre 2018 e 2020, que analisaram a aplicabilidade clínica de modelos de IA para diagnóstico em imagens 2D e 3D de estruturas dentárias e do complexo maxilofacial de indivíduos adultos. Os estudos analisados tiveram como foco a avaliação de redes neurais convolucionais na detecção automática de pontos cefalométricos, de lesões de cárie, de lesões apicais, de perda óssea periodontal, de sistemas de implantes, de cistos odontogênicos e tumores em maxila e mandíbula, de osteoporose, de sinusite maxilar, de terceiros molares e canal mandibular, de ateromas em carótida, de fratura radicular vertical, de osteoartrite em ATM, avaliação de morfologia radicular e numeração de elementos dentários. O desempenho dos modelos variou entre os diferentes algoritmos. Observou-se que 65% dos trabalhos analisados mostrou acurácia diagnóstica dos modelos acima de 80%. Os modelos propostos nos estudos mostraram potencial para utilização como auxiliares de diagnóstico por imagem nas diversas propostas analisadas. Contudo, estudos adicionais são necessários a fim de melhor investigar a confiabilidade destes modelos e de minimizar ou contornar os limites apresentados nos estudos, de forma a possibilitar a generalização de sua aplicabilidade clínica.

Palavras-chave: Aprendizado de Máquina; Inteligência Artificial; Rede Neural; Rede Neural Convolucional; Aprendizado Profundo; Algoritmo; Radiografia; Tomografia Computadorizada Cone Beam; Cefalometria; Radiografia panorâmica; Radiografia intraoral; Radiografia periapical; Detecção; Diagnóstico; Classificação; Automático; Ortodontia; Cárie Dental; Doença endodôntica; Lesão periapical; Periodontite; Canal Radicular; Polpa Dental; Odontogênico; Cistos; Tumores; Articulação temporomandibular; Mandíbula; Maxila; Odontologia; Seio maxilar; Perda óssea.

ABSTRACT

The present literature review aimed to investigate the literature related to the application of Artificial Intelligence (AI) in the analysis of imaging exams in various dental specialties, their performance in this task, possibilities of application in clinical practice, as well as their advantages and limitations. The bibliographic search conducted in a data source resulted in the inclusion of 58 original full-text articles, experimental papers and systematic literature reviews, published in English, a period of publication between 2018 and 2020, which analyzed the clinical applicability of AI models for diagnosis on 2D and 3D images of dental structures and maxillofacial complex of adult individuals. The studies analyzed focused on the evaluation of convolutional neural networks in the automatic detection of cephalometric points, caries lesions, apical lesions, periodontal bone loss, implant systems, odontogenic cysts and tumors in the maxilla and mandible, osteoporosis in mandible, maxillary sinusitis, third molars and mandibular canal, atherosclerotic carotid plaque, of vertical root fracture, osteoarthritis in TMJ, evaluation of root morphology and numbering of dental elements. The performance of the models varied between the different algorithms. It was observed that 65% of the analyzed studies showed diagnostic accuracy of the models above 80%. The models proposed in the studies showed potential for use as diagnostic imaging aids in the various proposals analyzed. However, further studies are needed to better investigate the reliability of these models and to minimize or circumvent the limits presented in the studies, in order to enable the generalization of their clinical applicability.

Keywords: Machine Learning; Artificial Intelligence; Neural Network; Convolutional Neural Network; Deep Learning; Algorithm; Radiography; Cone Beam Computed Tomography; Cephalometrics; X-ray; Panoramic Radiograph; Intraoral Radiograph; Periapical Radiograph; Detection; Diagnosis; Classification; Automated; Automatic; Orthodontics; Dental Caries; Endodontic Disease; Periapical Disease; Periapical Lesion; Periodontitis; Root Canal; Dental Pulp; Odontogenic; Cysts; Tumors; Temporomandibular Joint; Mandible; Maxilla; Dentistry; Sinus; Bone Loss.

LISTA DE ABREVIATURAS

2D	Duas dimensões, bidimensional
3D	Três dimensões, tridimensional
ATM	Articulação temporomandibular
AUC	<i>Area under the curve</i>
BL	<i>Bone level</i>
BRSI	Índice de severidade de reabsorção óssea linear
CAD	<i>Computer-aided diagnosis system</i>
CAM	<i>Class activation map</i>
CBCT	<i>Cone beam computed tomography</i>
CE	<i>European Conformity committee</i>
CNN	<i>Convolutional neural network</i>
DL	<i>Deep learning</i>
FDA	<i>Food and Drug Administration</i>
H	Hora(s)
IA	Inteligência artificial
ICC	Coeficiente de Correlação intraclasse
Min	Minuto(s)
ML	<i>Machine learning</i>
PCC	Coeficiente de Correlação Pearson
TC	Tomografia computadorizada
TCFC	Tomografia computadorizada de feixe cônico
TL	<i>Tissue level</i>

SUMÁRIO

1 INTRODUÇÃO	9
2 MATERIAIS E MÉTODOS	13
2.1 QUESTÕES DA REVISÃO.....	13
2.2 ESPECIALIDADES ODONTOLÓGICAS INCLUÍDAS NA BUSCA BIBLIOGRÁFICA.....	13
2.3 ESTRATÉGIA DE BUSCA	13
2.4 CRITÉRIOS DE INCLUSÃO.....	14
3 RESULTADOS	15
4 DISCUSSÃO	18
5 CONSIDERAÇÕES FINAIS	48
REFERÊNCIAS	50
ANEXO I	55
ANEXO II	58
ANEXO III	59
ANEXO IV	61
ANEXO V	63
ANEXO VI	64
ANEXO VII	66
ANEXO VIII	68
ANEXO IX	70
ANEXO X	71

1 INTRODUÇÃO

Inteligência artificial (IA) é um ramo da ciência da computação dedicada ao desenvolvimento de algoritmos para o cumprimento de tarefas tradicionalmente associadas à inteligência humana, como a habilidade de aprender e de resolver problemas (TANG *et al.*, 2018). Também pode ser definida como a capacidade de uma máquina de imitar a inteligência humana no desempenho de tarefas complexas, como resolução de problemas, reconhecimento de objetos e de palavras e tomada de decisão (HUNG *et al.*, 2020).

As tecnologias de inteligência artificial têm apresentado nos últimos anos um crescente sucesso e permeado um número cada vez maior de atividades rotineiras nos âmbitos pessoal e profissional, nas quais assumem papel auxiliar importante. No campo das atividades profissionais, a Odontologia tem se beneficiado dos modelos de IA em suas diversas áreas (HWANG *et al.*, 2019; PARK, W.J.; PARK, J.B., 2018; HUNG *et al.*, 2020), utilizando principalmente o aprendizado de máquina (*machine learning* – ML) e, em especial, o aprendizado profundo (*deep learning* – DL), que são subcampos da IA.

Aprendizado de máquina (*machine learning* – ML) é uma técnica para análise de dados que permite ao sistema computacional aprender a reconhecer padrões a partir de um grande conjunto de dados (CHEN *et al.*, 2020), tornando possível tarefas de classificação e a predição de resultados. Baseia-se em uma vasta variedade de algoritmos, comumente referidos como *modelos*. A escolha de um modelo em particular é determinada pelas características dos dados, bem como do tipo de resultado pretendido (NICHOLS *et al.*, 2019). Os dados utilizados podem ser, por exemplo, um conjunto de imagens. Estas imagens servirão como dados de *input* para o aprendizado do modelo, que envolve várias formas de reconhecimento de padrões, gerando um *output* na forma de classificação ou de identificação de uma determinada característica.

Nesse sentido, a Radiologia parece oferecer acesso mais direto ao uso de modelos de IA, já que produz imagens digitalmente codificadas que podem ser mais facilmente traduzidas para a linguagem computacional (HUNG *et al.*, 2020; LEITE *et al.*, 2020). Uma radiografia odontológica digital consiste em milhares de pixels, onde cada unidade de pixel representa diferentes níveis de brilho dentro da rede. O pixel

representando radiopacidade pode ser identificado como metal ou outra estrutura com alta densidade. Com base nestas e em outras características, programas de IA “aprendem” a analisar a imagem digital durante a fase de treinamento, que envolve coordenação entre um engenheiro de *software* e um especialista clínico (CHEN *et al.*, 2020). Partindo de uma grande quantidade de radiografias, o especialista seleciona algumas para compor o conjunto de treinamento e as analisa manualmente. O engenheiro de *software* utiliza esse conjunto de treinamento para arquitetar o modelo matemático de aprendizagem para ensinar o programa de IA a habilidade de fazer generalizações e continuar reconhecendo automaticamente, na amostra restante, as informações identificadas pelo especialista.

Dessa forma, os algoritmos do ML evoluem conforme são expostos a mais dados e, à medida que analisam mais informações dos pixels de centenas ou milhares de radiografias, “aprendem” a fornecer respostas específicas, em consonância com os parâmetros definidos manualmente pelo especialista (TANG *et al.*, 2018). Dessa forma, pode-se dizer que o acréscimo de dados ou informações de treinamento geralmente acarreta melhora do desempenho do modelo.

Aprendizado Profundo (*Deep Learning* – DL), por sua vez, é um tipo de ML, em que o conhecimento é representado por uma hierarquia de conceitos interligados, compondo diversas “camadas”, nas chamadas “redes neurais” (LEITE *et al.*, 2020). A ideia de redes neurais surgiu como uma tentativa de imitar os padrões de conexão dos neurônios no córtex visual animal (LEE *et al.*, 2018). Na rede neural, vários “neurônios” são conectados entre si de modo a formar uma rede, formada por três tipos de camadas, chamadas de camada de entrada (*input*), camada de saída (*output*) e camadas ocultas (*hidden*). A camada de entrada recebe os dados de entrada e o valor final é gerado na camada de saída. As camadas ocultas fazem cálculos intermediários que auxiliam a rede a encontrar os valores finais. Este modelo, conforme for alimentado por quantidades massivas de dados, sofre uma automelhoria e um autorrefinamento constantes. Isso ocorre quando o modelo tem acesso a um conjunto de dados de treinamento, em que os processos de aprendizado são repetidos automaticamente, sem necessidade de interferência manual para definição de características da imagem analisada. Assim, os sistemas de aprendizado profundo são capazes de aprender de forma autônoma com as características detectadas na imagem e, simultaneamente, desempenhar a classificação dessa imagem (HIRAIWA

et al., 2019). Além disso, os modelos de aprendizado profundo conseguem detectar variações anatômicas pronunciadas e apresentam baixa sensibilidade a artefatos nas imagens (DOT *et al.*, 2020).

Há diversos tipos de redes neurais; contudo, a chamada rede neural convolucional (CNN - *convolutional neural network*) é um tipo de algoritmo de DL especializado em lidar com dados de topologia tipo grade, como imagens 2D e 3D, sendo, por esta razão, aplicado à análise e classificação de imagens para fins diagnósticos. As várias camadas nestes algoritmos são usadas para detectar características que variam de simples, como linhas, bordas, texturas e intensidade, até complexas, como formas, lesões ou órgãos em sua totalidade (TANG *et al.*, 2018).

O principal constituinte de uma rede neural é o chamado neurônio artificial, que é um modelo matemático não-linear inspirado no neurônio humano (SCHWENDICKE *et al.*, 2020). Uma CNN apresenta arquitetura análoga ao padrão de conectividade dos neurônios no cérebro humano e requer pré-processamento mínimo, “aprendendo” sozinha os parâmetros que um algoritmo tradicional necessita que sejam implementados manualmente, como acontece em outras técnicas de ML. Neste tipo de rede neural, cada camada contribui com uma pequena informação, reconhecida e assimilada por convolução pela camada seguinte, até chegar à chamada camada *pooling*, que faz a classificação final do objeto (NICHOLS *et al.*, 2019). Para chegar até este estágio, é necessário um período de treinamento do modelo. Quando este período é finalizado, a rede neural é capaz de usar informações para inferir as regras, com o aprendizado e o mapeamento das características sendo feitos ao mesmo tempo (SCHWENDICKE *et al.*, 2020). Nesse ponto, a efetividade do modelo precisa ser avaliada através dos dados de teste e vai depender de diversos fatores, tais como o tipo e quantidade de dados usados para treinamento, o tipo de CNN utilizado e o número de camadas existentes no modelo.

A aplicabilidade clínica do DL se refere principalmente à análise de imagens, com tarefas dentro de três categorias: detecção, para identificar anormalidade em uma imagem; segmentação, em que uma estrutura de interesse é isolada, como na definição dos limites de um órgão na imagem; e classificação, em que uma característica ou lesão na imagem é definida dentro de uma categoria (TANG *et al.*, 2018; NICHOLS *et al.*, 2019). A aplicabilidade do DL mostra, assim, potencial para automatizar tarefas que normalmente requerem muito tempo para serem executadas

e para melhorar a capacidade humana de lidar com dados massivos e complexos. De fato, nas próximas décadas, a tendência parece ser a automatização de tarefas puramente manuais como consequência do progresso tecnológico (GRISCHKE *et al.*, 2020).

No âmbito da radiologia odontológica, a IA parece ocupar um espaço crescente em estudos que buscam avaliar sua aplicabilidade clínica em diversas especialidades como método auxiliar ao radiologista, com acurácia diagnóstica próxima daquela dos especialistas humanos (HWANG *et al.*, 2019; LEITE *et al.*, 2020), de forma a otimizar o fluxo de trabalho com qualidade e segurança. Nesse sentido, a Associação Canadense de Radiologistas (CAR), a respeito do uso da IA na Radiologia, recomenda que a comunidade de radiologistas deva se familiarizar com as diferentes técnicas de IA e se capacitar a analisar criticamente as oportunidades e desafios associados à introdução de novas ferramentas de IA (TANG *et al.*, 2018). Da mesma forma, o laboratório de IA (AI-LAB) do Colégio Americano de Radiologia (ACR) oferece ferramentas para radiologistas projetadas para ajudá-los a aprender o básico da IA e participar diretamente da criação, validação e uso da IA para cuidados de saúde.

Paralelamente à possibilidade de modelos de IA se provarem preciosos auxiliares nas tarefas rotineiras do dentista radiologista, dúvidas necessariamente se interpõem neste cenário, referentes principalmente ao desempenho destes modelos, sua confiabilidade, e à uma possível futura redefinição do papel dos radiologistas enquanto *experts* da análise de imagens.

Assim, a presente revisão tem como objetivo investigar a literatura relacionada à aplicação da Inteligência Artificial na análise de exames de imagem nas diversas especialidades odontológicas, seu desempenho nesta tarefa, possibilidades de aplicação na prática clínica, bem como suas vantagens e limitações.

2 MATERIAIS E MÉTODOS

2.1 QUESTÕES DA REVISÃO

- Qual o método de Aprendizado Profundo predominante na análise de imagens radiográficas odontológicas pelos sistemas de IA avaliados, e por quê?

- Quais os resultados mostrados, no que se refere ao desempenho e confiabilidade dos sistemas avaliados?

- Quais os benefícios ou vantagens que a utilização da IA pode trazer, quando aplicada aos exames de imagem odontológicos?

- Quais as possíveis limitações relacionadas à utilização da IA, quando aplicada aos exames de imagem odontológicos?

2.2 ESPECIALIDADES ODONTOLÓGICAS INCLUÍDAS NA BUSCA BIBLIOGRÁFICA

Considerando o grande volume de publicações no período abrangido pela busca, optou-se por selecionar aquelas referentes às áreas que possivelmente representam maior interesse para a Radiologia Odontológica, a saber: Cariologia, Endodontia, Periodontia, Ortodontia, Implantodontia e Patologia Bucal.

2.3 ESTRATÉGIA DE BUSCA

Foram realizadas buscas bibliográficas no período de janeiro a dezembro de 2020, por trabalhos de texto integral na base de dados MEDLINE/PubMed.

A fim de obter uma visão ampla da aplicabilidade dos sistemas de IA na análise de imagens odontológicas, foram utilizadas como palavras-chave combinações de termos do *Medical Subject Headings* (MeSH), com a expressão booleana “NOT” para excluir resultados indesejáveis, como estudos utilizando dados genômicos. As palavras-chave utilizadas podem ser divididas em categorias, como segue:

a) sobre aprendizado de máquina: *machine learning* (ML); *artificial intelligence* (AI); *neural network*; *convolutional neural network* (CNN); *deep learning*; *algorithm*;

b) sobre métodos de imagem: *radiography; cone beam computed tomography (CBCT); cephalometrics; X-ray, panoramic radiograph; intraoral radiograph; periapical radiograph;*

c) sobre métodos de processamento de aprendizado de máquina: *detection; diagnosis; classification; automated; automatic;*

d) sobre aplicabilidade na odontologia: *orthodontics; dental caries; endodontic disease; periapical disease; periapical lesion; periodontitis; root canal; dental Pulp; odontogenic; cysts; tumors; temporomandibular joint; mandible; maxilla; dentistry; sinus; bone loss.*

2.4 CRITÉRIOS DE INCLUSÃO

Foram incluídos artigos originais de texto completo, trabalhos experimentais e revisões sistemáticas da literatura, publicados em inglês, período de publicação entre 2018 e 2020, que utilizassem, em sua metodologia, exames de imagem com radiação ionizante, como tomografia computadorizada, radiografias panorâmicas, periapicais e interproximais do complexo dentomaxilofacial de sujeitos humanos, exames estes submetidos a modelos de diagnóstico baseados em algoritmos de inteligência artificial. O foco do trabalho deveria ser a aplicação de modelo de IA no diagnóstico ou detecção de condição patológica ou estruturas anatômicas relevantes para a área clínica em questão.

Foram encontrados 486 artigos e, destes, selecionados 58, de forma a atender os critérios de inclusão. Estes 58 artigos analisaram a aplicabilidade clínica de modelos de IA para diagnóstico em imagens 2D e 3D de estruturas dentárias e do complexo maxilofacial de indivíduos adultos, nas diversas áreas da Odontologia.

3 RESULTADOS

Esta revisão da literatura incluiu 58 artigos publicados entre 2018 e 2020, abrangendo a aplicabilidade da IA na análise de radiografias em diversas especialidades odontológicas, parecendo haver uma tendência de crescimento, ao longo do tempo, na publicação deste tipo de pesquisa nas fontes de busca. Deste total, 49 artigos apresentaram trabalhos experimentais que testaram o desempenho de modelos de IA em tarefas de identificação, classificação e segmentação de estruturas em imagens radiográficas. De acordo com a busca realizada, a área de maior volume de publicações foi a Radiologia, com dezesseis artigos.

A maioria dos trabalhos utilizou *Deep Learning*, especialmente redes neurais convolucionais, como método específico de IA, demonstrando que este método está associado a um campo de pesquisa altamente dinâmico, sendo considerado o mais promissor na Odontologia, especialmente por suas vantagens em relação a outras modalidades de IA. De fato, o método de *Deep Learning* consegue lidar com variações anatômicas pronunciadas, além de apresentar baixa sensibilidade a artefatos na imagem (DOT *et al.*, 2020).

A maioria dos trabalhos foi de avaliação de acurácia diagnóstica de modelos de CNN desenvolvidos para o estudo, aperfeiçoados ou já existentes. Os estudos abordaram problemas clínicos nas seguintes áreas: Cariologia (cinco trabalhos), Endodontia (seis trabalhos), Implantodontia (cinco trabalhos), Periodontia (cinco trabalhos), Ortodontia (dez trabalhos), Patologia (seis trabalhos) e aplicações pertinentes principalmente à Radiologia Odontológica (16 trabalhos).

Os conjuntos de dados utilizados nos estudos foram muito variáveis em tamanho, desde algumas dezenas de imagens a mais de onze mil imagens. A modalidade de imagem variou conforme a aplicabilidade na área em questão. A modalidade mais utilizada foi a radiografia panorâmica (26 estudos), seguida de TCFC (11 estudos), periapicais (nove estudos), radiografias cefalométricas (três estudos), e Waters, interproximal e TC convencional, com um estudo cada. A origem das amostras utilizadas foi, de modo geral, bancos de dados e registros médicos de Instituições de Ensino e de clínicas privadas, bem como, no caso de dois estudos (MONTÚFAR *et al.*, 2018a; MONTÚFAR *et al.*, 2018b), um repositório público virtual de imagens ligado à Instituição de ensino.

Dentre as publicações na área da Ortodontia, quando do momento da busca, as de maior ocorrência foram as relacionadas à testagem de modelos para automatização de localização de pontos cefalométricos. Por este motivo, e dada a importância desta etapa no tratamento ortodôntico, estas foram as selecionadas para a presente revisão. Os resultados observados nos nove trabalhos clínicos selecionados que tratam da aplicação da IA às imagens na Ortodontia foram sumarizados na Tabela I (Anexo I).

A heterogeneidade de metodologias utilizadas nos trabalhos não possibilita uma comparação direta entre resultados, mas, levando-se em conta aqueles que utilizaram a mesma modalidade de imagens (tridimensionais), padrões de referência semelhantes e cujos resultados foram mensurados em termos de erro médio global na localização dos pontos cefalométricos obtidos, pode-se estabelecer comparação de valores desta média de erro, conforme Tabela II (Anexo II).

Na área de Cariologia, a utilização de modelos de IA concentrou-se na análise de radiografias para detecção de lesões de cárie, e os resultados obtidos nos quatro estudos clínicos selecionados estão sumarizados na Tabela III (Anexo III).

Dos cinco trabalhos experimentais relacionados à Endodontia, quatro abordaram a utilização de modelos de IA na detecção de lesões radiolúcidas periapicais, e um buscou analisar a morfologia radicular de molares inferiores. Os resultados destes estudos podem ser observados nas tabelas IV (Anexo IV) e V (Anexo V).

A busca bibliográfica realizada na área de Periodontia retornou cinco estudos, selecionados para esta revisão. Todos tiveram como objeto a detecção automática, em imagens radiográficas, de dentes com doença periodontal, utilizando CNN. Os resultados destes estudos podem ser observados na tabela VI (Anexo VI).

Foram incluídos cinco estudos na área de Implantodontia, todos do ano de 2020, que focaram na identificação automática de sistemas de implantes em radiografias panorâmicas ou periapicais, e cujos resultados podem ser vistos na tabela VII (Anexo VII).

Abrangendo a área de Patologia Bucal, foram incluídos seis artigos, cinco deles experimentais, testando CNN na identificação automática de cistos odontogênicos e tumores em panorâmicas e imagens de TCFC, conforme visto nas tabelas VIII e IX (Anexos VIII e IX).

Foram incluídos nesta revisão, ainda, dezesseis estudos rotulados como relacionados à Radiologia Odontológica, por abordarem aspectos de diagnóstico por imagens geralmente afeitos à prática desta especialidade, o que claramente não exclui sua aplicabilidade para dentistas de outras áreas clínicas. O resumo dos resultados e conclusões destes estudos podem ser vistos na tabela X (Anexo X).

O referencial comparativo para análise do desempenho dos modelos testados foi, em geral, a análise das imagens feita por especialistas humanos (31 trabalhos), a comparação com outros exames de imagem ou histopatológico (oito trabalhos), comparação com outros modelos de IA (seis trabalhos) e, em três trabalhos, não ficou claro ou não houve referencial comparativo.

O desempenho dos modelos variou entre os estudos conforme a tarefa a ser desempenhada, o tipo de modelo utilizado, o número de camadas das CNN, a quantidade da amostra de treinamento, o referencial comparativo, a modalidade de imagem utilizada, se houve pré-tratamento das imagens, e os parâmetros de medição utilizados. A acurácia diagnóstica foi maior do que 80% em 65% dos estudos. Seis trabalhos consideraram claramente que o desempenho do modelo testado foi superior ao dos especialistas humanos, e 11 estudos consideraram este desempenho similar ao dos especialistas.

4 DISCUSSÃO

Ao realizar a análise dos artigos relacionados à Ortodontia incluídos nesta revisão, observa-se que os esforços foram direcionados para o desenvolvimento de algoritmos que automatizassem a tarefa de localização dos pontos cefalométricos, priorizando-se os exames de imagem tridimensionais (tomografias computadorizadas).

Sabe-se que a marcação dos pontos cefalométricos constituem etapa vital no planejamento ortodôntico, sendo geralmente realizada via identificação manual, na imagem, de pontos que representam características morfológicas craniofaciais, requerendo, assim, alto nível de experiência, de tempo e de trabalho, mesmo para especialistas. Portanto, há uma demanda crescente por um sistema de referência automático auxiliado por computador que possa melhorar o fluxo de trabalho e agilizar esta etapa.

Além disso, os recentes avanços em tecnologias de imagem facilitaram uma mudança da cefalometria 2D para 3D, trazendo várias vantagens, como possibilitar identificação mais precisa de estruturas anatômicas e mínima distorção geométrica nas imagens (YUN *et al.*, 2020). Assim, a tomografia computadorizada é considerada, por alguns autores, como tendência de técnica de imagem na ortodontia (MONTÚFAR *et al.*, 2018a). Dos nove trabalhos analisados, seis adotaram como modalidade de imagem de estudo a tomografia computadorizada (Anexo I, Tabela I).

Os estudos revisados, em sua totalidade, propuseram um método ou modelo de IA a ser aplicado na detecção de um número determinado de pontos cefalométricos nas imagens da amostra, de forma a automatizar e, conseqüentemente, agilizar o processo. O desempenho desse modelo foi avaliado comparando-se a discrepância posicional entre os resultados obtidos e os dos especialistas, numa medida do erro médio, para cada ponto, em milímetros. Ao final, foi calculado o índice de erro médio global. Apenas no estudo de Park *et al.* (2019), um modelo de IA foi comparado a outro modelo de IA. Cabe ressaltar que, para fins de aplicabilidade clínica do modelo proposto, considera-se que este índice deva ser igual ou menor que 2mm.

Apenas um estudo com modalidade de imagem tridimensional (NEEPALU *et al.*, 2018) obteve índice dentro do parâmetro aceitável, com erro médio global igual a 1,88mm. Os estudos com imagens bidimensionais mostraram melhores resultados de

localização, com índice de erro menor do que 0,2mm para parâmetros lineares (KUNZ *et al.*, 2020) e menor do que 0.9mm (HWANG *et al.*, 2020).

Estes resultados podem ser devidos, em grande parte, à grande variação anatômica individual das imagens das amostras, e ao grande número de conjunto de dados necessários para o treinamento do modelo, dificilmente acessível aos pesquisadores. Embora os hospitais e algumas clínicas possam gerar muitos conjuntos de dados de tomografia computadorizada, poucos destes dados podem ser usados para pesquisas, por razões éticas e legais.

Considerando este empecilho em particular, YUN *et al.* (2020) desenvolveram modelo de IA com estrutura de aprendizagem multiestágio, que objetiva lidar com o desafio de uma pequena quantidade de dados ao aprender recursos 3D a partir de dados de TC 3D. Contudo, estudos adicionais nesse sentido ainda são necessários, a fim de que se possa alcançar resultados que corroborem o uso do modelo na clínica.

Alguns dos estudos mostraram dificuldade do modelo na localização de alguns dos pontos cefalométricos, com maior incidência de erro para os forames infra e supraorbitais (ED-DHAHRAOUY *et al.*, 2018), porio e sela (MONTÚFAR *et al.*, 2018a) e gônio, condílio, R1 e sela (NEELAPU *et al.*, 2018). As possíveis causas apontadas para esta falta de precisão seriam a baixa definição do volume, variação anatômica e formato irregular das estruturas, o que dificultaria seu reconhecimento por parte da IA e requereria ajustes manuais. Em se tratando de um ponto importante do desempenho dos modelos, nota-se, por parte dos autores, preocupação em sobrepujar esta dificuldade. Montúfar *et al.* (2018b) criaram uma técnica que permite recortar subvolumes da imagem para ajuste mais preciso do ponto por parte do modelo, diminuindo significativamente, assim, o erro médio global na localização dos pontos. Chen *et al.* (2020) desenvolveram técnica de autossegmentação por reconhecimento automático dos limites anatômicos do volume de interesse, inclusive em imagens de baixa qualidade, sem necessidade de pós processamento manual, obtendo pequena diferença (2 voxels) entre os grupos teste e controle no plano sagital mediano.

O número de pontos cefalométricos para localização teve grande variação nos trabalhos analisados, desde três até 93. Contudo, conforme aponta Dot (2020), até o momento não há uma lista de pontos cefalométricos para análise tridimensional reconhecida mundialmente, e a maioria apresentou uma adaptação de técnicas bidimensionais pré-existentes.

Quanto à possibilidade de agilizar o fluxo de trabalho, alguns dos estudos indicam que o uso dos modelos de IA realmente podem tornar esta tarefa mais rápida, com tempos de processamento de cada conjunto de imagens variando de poucos segundos (MONTÚFAR *et al.*, 2018b; KUNZ *et al.*, 2020) a quinze minutos para a segmentação das estruturas (CHEN *et al.*, 2020).

Portanto, em se tratando de análise cefalométrica automatizada em imagens 3D, os métodos propostos apresentaram desempenho promissor, sugerindo um possível avanço da cefalometria para uma fase tridimensional, embora, presentemente, ainda não atenda às exigências para aplicação clínica imediata.

Para as imagens 2D, houve acurácia na localização dos pontos e redução de tempo de trabalho, quando comparado ao especialista humano, sugerindo resultados compatíveis com parâmetros para utilização clínica, dentro das limitações inerentes a estudos *in vitro* (PARK *et al.*, 2019; KUNZ *et al.*, 2020).

A utilização de modelos de IA na área de Cariologia, aplicados à análise de imagens radiográficas, concentrou-se basicamente na tentativa de identificação de lesões de cárie, em diversos estágios, a fim de proporcionar ao clínico uma ferramenta que agilize o processo de diagnóstico e que, acima de tudo, torne esse processo mais confiável. De fato, a detecção rápida e precisa é fundamental na implementação de prevenção e tratamento adequados em pacientes com cárie dentária (LEE *et al.*, 2018).

Tradicionalmente, o diagnóstico de lesões de cárie baseia-se nos exames visual/tátil e radiográfico (GEETHA *et al.*, 2020; PRADOS-PRIVADO *et al.*, 2020). A radiografia, juntamente aos exames clínicos, desempenha um papel importante no diagnóstico, principalmente nos casos de lesões proximais e/ou radiculares. No entanto, estudos relataram grandes variações na confiabilidade e precisão de detecção, de acordo com a metodologia do estudo e o nível de experiência do clínico (LEE *et al.*, 2018; CANTU *et al.*, 2020). A cárie não diagnosticada adequada e oportunamente pode acarretar desde um subtratamento, muitas vezes culminando com a perda do dente, até o sobretratamento, com perda desnecessária de estrutura dentária em tratamentos invasivos e maiores custos de tratamento.

Nesse contexto, a auxílio de um método automatizado para detecção de cáries nas radiografias representaria um ganho significativo para o paciente e para o profissional. Esse é um ponto importante tanto para o radiologista, que na maioria das

vezes não dispõe de dados clínicos para o diagnóstico, baseando-se unicamente na imagem analisada, quanto para o clínico, que tem na radiografia um valioso meio complementar, principalmente nos casos em que os exames clínicos se mostram inconclusivos.

Comercialmente, o único *software* criado especificamente para o diagnóstico de cáries foi o *Logicon Caries Detector*TM, da Carestream Dental (LCDS, Carestream Dental, GA, EUA). Contudo, ainda permanece controverso se este sistema seria mais preciso do que examinadores clínicos experientes (LEE *et al.*, 2018).

Outro sistema disponível comercialmente e que, dentre suas funcionalidades, poderia detectar lesões de cárie, é o *Diagnocat* (Diagnocat Inc, CA, EUA). Contudo, a busca realizada para esta revisão não encontrou nenhum resultado contemplando este *software*, no que se refere a diagnóstico de lesões de cárie. Segundo a página do *Diagnocat* na internet (diagnocat.com), o sistema está em processo de obtenção de licença pela *Food and Drug Administration* (FDA), agência federal dos Estados Unidos.

Os resultados dos trabalhos analisados (Anexo III, Tabela III) mostram que a utilização dos modelos de IA na análise de radiografias para detecção de lesões de cárie pode ser considerada promissora, com precisão e eficiência consideráveis, refletidos nos índices de acurácia de 80%, aproximadamente. Apesar disso, e talvez justificando, em parte, os índices encontrados, os estudos se depararam com algumas limitações importantes, relacionadas basicamente com a metodologia de IA utilizada, tais como: a aplicação de uma metodologia de *Deep Learning* requer uma amostra de treinamento muito grande, e um longo tempo de treinamento; os parâmetros da rede neural têm que ser determinados empiricamente e a fase de treinamento requer pesado investimento em sistemas computacionais. Estes fatores certamente tendem a aumentar a variabilidade dos resultados, dificultando a transposição destes para a situação de prática clínica. O estudo de Geetha *et al.* (2020) obteve o mais alto índice de acurácia na detecção de lesões de cárie, sobressaindo em relação aos demais. Ressalta-se, no entanto, que a rede neural utilizada continha apenas uma camada, e a amostragem foi a menor entre os trabalhos (n=105).

Também houve variabilidade na modalidade de radiografia utilizada. Observou-se que dois estudos utilizaram radiografias periapicais (LEE *et al.*, 2018; GEETHA *et al.*, 2020), um estudo utilizou radiografias interproximais (CANTU *et al.*, 2020) e um

estudo não especificou o tipo de radiografia, apenas mencionou serem de sistema digital. Os estudos não justificaram a escolha do tipo de radiografia utilizada. Sabe-se que as radiografias interproximais são a modalidade de imagem de eleição para diagnóstico de lesões de cárie. Assim, uma possível justificativa para a utilização de radiografias periapicais neste tipo de estudo poderia ser, mais uma vez, a dificuldade de obter uma amostragem quantitativamente significativa, esbarrando-se em questões éticas e legais no acesso aos bancos de imagens das instituições. Esta questão torna-se relevante no que se refere aos resultados, na medida que as radiografias periapicais podem dificultar a visualização de lesões de cárie iniciais, restritas a esmalte, o que poderia influenciar negativamente os índices de acurácia e de sensibilidade obtidos.

Quanto aos parâmetros comparativos dos resultados de cada estudo, apenas um estudo comparou os resultados do modelo de IA com diagnósticos realizados por especialistas (CANTU *et al.*, 2020). Esta seria a situação que melhor refletiria o valor da aplicabilidade clínica do modelo pois, para o profissional da clínica odontológica, o modelo de IA deveria ser transformado em um *software* cujo desempenho fosse superior ao seu em tempo de processamento e confiabilidade dos resultados.

O estudo de Cantu *et al.* (2020) também foi o único a diferenciar a profundidade das lesões de cárie detectadas, verificando que a rede neural testada mostrou sensibilidade acima de 0,70 para lesões iniciais e avançadas, e que os dentistas em grande parte apresentaram baixa sensibilidade para lesões iniciais: todos, exceto um, mostraram sensibilidade abaixo de 0,25. Para lesões avançadas, os índices de sensibilidade ficaram entre 0,40 e 0,75. Os autores consideram que, no caso do diagnóstico de lesões iniciais, provavelmente os dentistas foram mais prudentes em identificar alterações de imagem como lesões iniciais de cárie, temendo falsos positivos e sobretratamento, enquanto a CNN não mostrou tal comportamento.

Um fator importante a ser levado em conta em relação às radiografias de treinamento é que, quando analisadas por especialistas humanos, a definição dos critérios para a identificação da doença deve ser clara. Este é um parâmetro fundamental para a correta especificação das variáveis que a rede neural tem a aprender. Em outras palavras, concordância entre os examinadores é requisito para resultados mais confiáveis (PRADOS-PRIVADO *et al.*, 2020). Dessa forma, o estudo

de Lee *et al.* (2018) foi o único a esclarecer este aspecto, o qual era parte essencial do estudo.

A grande variabilidade das metodologias utilizadas em cada estudo não permite comparações entre eles quanto ao desempenho diagnóstico das redes neurais testadas. De forma geral, e individualmente, mostraram resultados promissores no uso de modelos de IA na detecção de lesões de cárie em radiografias. Contudo, para fins de utilização efetiva na prática clínica, parece haver ainda um longo caminho a ser percorrido, no que diz respeito à confiabilidade e previsibilidade de resultados.

Na área de Endodontia, no período pesquisado, os estudos tiveram como foco principal a detecção de lesões radiolúcidas periapicais pelos modelos de IA (Anexo IV, Tabela IV). Apenas um dos cinco trabalhos incluídos na presente revisão direcionou-se de forma diferente, testando a utilização da IA para analisar a morfologia radicular dentária (HIRAIWA *et al.*, 2019).

Os índices de detecção de lesões radiolúcidas periapicais nas imagens radiográficas apresentam grande variabilidade interexaminadores. As variações de interpretação entre examinadores e seus vieses poderiam ser reduzidas com o uso de sistemas de IA. Ekert *et al.* (2019) consideram que a possibilidade de aplicar CNNs nesta tarefa pode aumentar significativamente a acurácia do diagnóstico, tornando-a semelhante ou superior à de especialistas experientes, além de poupar tempo e poder semiautomatizar o processo de documentação. Assim, testaram o desempenho de um modelo CNN de aprendizado profundo na identificação de lesões radiolúcidas periapicais em radiografias panorâmicas segmentadas, comparando-o com detecção por seis examinadores experientes independentes. Os resultados mostraram que, quando a taxa de concordância geral entre examinadores foi de 54,6%, o índice AUC da CNN foi de 0.85. Em caso de 100% de concordância dos examinadores, o índice AUC subiu para 0.95. O modelo, de forma geral, apresentou sensibilidade limitada devido à baixa prevalência de lesões nas amostras, necessitando melhoria antes da etapa de aplicação clínica.

No estudo de Endres *et al.* (2020), para detecção de lesões periapicais, houve a preocupação em justificar a escolha da radiografia panorâmica como modalidade de imagem, já que as periapicais seriam a opção tradicional para avaliação de lesões periapicais. Os autores alegaram ter acesso a grande número de panorâmicas na Instituição de trabalho, visto serem os exames padrão de rotina devido à sua boa

habilidade discriminatória diagnóstica geral, além da baixa dosagem de radiação. Além disso, foi feita a discriminação entre as várias radiolucências periapicais que poderiam ser diagnosticadas, tais como lesão periapical inflamatória, granuloma periapical, cisto periapical, osteomielite e tumor. A metodologia também incluiu verificação cruzada com dados clínicos para estabelecimento do padrão ouro. Dessa forma, nesse grupo, a detecção das lesões baseou-se também em dados clínicos, além das imagens radiográficas, e não apenas na concordância interexaminadores, acarretando a provável diminuição da ocorrência de vieses e limitação da validação do estudo. Quanto aos resultados, estes foram semelhantes aos encontrados no estudo de Ekert *et al.* (2019). O desempenho do modelo de IA foi avaliado de acordo com diversos parâmetros, dentre eles o escore F1, medida de desempenho comum utilizada no aprendizado de máquina, representando a média harmônica da precisão e da sensibilidade do modelo. O modelo resultou em escore F1 de 0.58, com melhor desempenho diagnóstico que 58% dos dentistas, apresentando potencial como ferramenta complementar de diagnóstico, segundo os autores. Eles ressaltam que tanto a FDA quanto o *European Conformity committee* (CE) permitem o uso deste tipo de *software* apenas como ferramenta auxiliar, até o momento.

Outros estudos (ORHAN *et al.*, 2020; SETZER *et al.*, 2020) consideraram a TCFC como modalidade de imagem mais adequada para testar o desempenho dos modelos na detecção de lesões radiolúcidas periapicais, já que normalmente permite melhores condições para visualização de alterações periapicais e ósseas quando comparada às imagens bidimensionais, inclusive das lesões em estágio inicial.

Orhan *et al.* (2020) avaliaram, em 153 TCFC, o desempenho diagnóstico de um sistema do tipo CNN (*Diagnocat*) na detecção automática de lesões apicais, com medição de volume dessas lesões. Os resultados foram comparados com a média dos valores de localização, feita manualmente, por três especialistas. O índice de acurácia do sistema testado foi de 92,8%, tendo detectado 142 de um total de 153 lesões. Quanto à medição do volume das lesões, os valores foram equivalentes à medição manual. Os autores concluíram que não houve diferença estatisticamente significativa entre o sistema CNN e os especialistas humanos quanto ao desempenho. Ainda, foram feitas observações sobre alguns fatores relacionados às estruturas adjacentes às lesões que podem influenciar a capacidade de análise e de reconhecimento do sistema. Estes incluíram a capacidade de diferenciar a densidade

de tecidos moles e a área da lesão, que ficam semelhantes em uma TCFC, já que o exame não apresenta boa resolução para tecidos moles. Alterações como lesões endoperio, defeitos periodontais e perda de osso alveolar também podem alterar as medidas realizadas pelo sistema. O mesmo pode acontecer com estruturas anatômicas vizinhas normais, como canal incisivo, canal mandibular, forame mental, seio maxilar e fossa nasal, ao serem segmentadas nas imagens, e anomalias dentais, como *dens in dente*, ou mesmo ápice radicular com formação incompleta. Os autores sugerem fazer posteriormente alterações de programação do modelo contemplando variações de anatomia, a fim de minimizar este problema.

Em outro trabalho utilizando TCFC e detecção automática de lesões periapicais por modelo de IA, a metodologia incluiu uma definição de lesão periapical a ser identificada por examinadores humanos e pelo modelo: haveria lesão periapical quando detectada interrupção da linha da lâmina dura, e a área hipodensa associada ao ápice radiográfico fosse pelo menos duas vezes mais profunda do que o espaço do ligamento periodontal (SETZER *et al.*, 2020). Neste estudo, foi utilizada uma amostra de teste de 20 imagens com 61 raízes, e o desempenho do modelo foi comparado com o de três examinadores. Os resultados obtidos mostraram sensibilidade de 0.93, especificidade de 0.88 e acurácia de 0.87. Estes índices foram considerados excelentes pelos autores. A quantidade da amostra não foi considerada uma limitação do estudo pelos autores, por acreditarem que o modelo de IA aplicado, com arquitetura U-Net, utiliza a técnica de *data augmentation*, aumentando artificialmente a quantidade das amostras.

No estudo de Hiraiwa *et al.* (2019), a utilização de modelos de IA teve como foco a avaliação da morfologia radicular de primeiros molares inferiores em radiografias panorâmicas, na busca pela raiz disto-lingual. Esta raiz está presente em 3% a 23% dos casos, conforme características raciais e étnicas, e sua identificação e correta abordagem no momento do tratamento endodôntico está diretamente relacionada ao sucesso ou insucesso do tratamento. Um canal não tratado por não ter sido identificado pode levar à colonização microbiana e resultar na falha do tratamento. Nesse sentido, as imagens de TCFC oferecem alto grau de acurácia, mas não são exames realizados rotineiramente na clínica endodôntica. Assim, os autores consideram que, se a morfologia radicular pudesse ser prevista antes do tratamento, em radiografias panorâmicas ou periapicais, seria possível identificar aqueles

pacientes que iriam requerer TCFC para o planejamento do caso. Caberia aqui, talvez, o questionamento da real necessidade de TCFC se a análise prévia das imagens bidimensionais por um modelo de IA já pudesse fornecer índices aceitáveis de acurácia, semelhantes ou superiores aos daqueles obtidos por radiologistas experientes, como no estudo em questão.

Novamente, os trabalhos se depararam com a limitação de uma amostra relativamente pequena na etapa de treinamento dos modelos, o que acarreta impacto direto no desempenho do modelo na fase de testes. Segundo Hirawa *et al.* (2019), o desempenho de um sistema de aprendizado profundo depende tanto da qualidade quanto da quantidade do conjunto de dados de treinamento. Visando contornar este problema, muitos trabalhos utilizam o artifício do *data augmentation*, ou aumento de dados, que consiste no aumento artificial do número de imagens através da alteração do brilho, contraste, rotação e nitidez das imagens do conjunto de treinamento original. Contudo, mesmo utilizando-se esse recurso, a amostra pode continuar insuficiente, principalmente se o estudo tiver como foco apenas uma população específica.

Em uma revisão sistemática da literatura, são mostradas as aplicações da IA na endodontia contidas em 10 artigos originais de pesquisa publicados de janeiro de 2000 a junho de 2020 (BOREAK, 2020). Essas aplicações incluíram avaliação da morfologia radicular, predição de patologia periapical, detecção e diagnóstico de fratura radicular vertical, determinação do comprimento de trabalho, localização do forame apical, e predições de retratamento. As duas primeiras foram abordadas na presente revisão, enquanto as demais não atenderam ao critério do período de publicação. O autor considera que os trabalhos revisados mostram desempenho semelhante das redes neurais ao dos especialistas, com melhores precisão e acurácia, e podem ser de grande ajuda como uma segunda opinião especializada para o clínico menos experiente ou não-especialista.

Na área de Periodontia, os cinco estudos revisados (Anexo VI, Tabela VI) tiveram como objeto avaliar o desempenho de CNNs na detecção automática de dentes com comprometimento periodontal em imagens radiográficas, visando reduzir o esforço do profissional e prover uma “segunda opinião” acerca do diagnóstico.

Lee *et al.* (2018) consideram que, apesar dos avanços nas modalidades de tratamento, não tem havido melhora significativa na metodologia para diagnóstico e predição da doença periodontal. Tanto o diagnóstico clínico quanto o prognóstico

dependem de evidências empíricas, já que depende da força de sondagem, angulação, localização e diâmetro da ponta da sonda (CHANG *et al.*, 2020). Além disso, o diagnóstico e intervenção precoces desempenhariam papel crucial no desfecho clínico da doença (KIM *et al.*, 2019), justificando a busca por um sistema automatizado de diagnóstico de doença periodontal em exames complementares de imagem.

Lee *et al.* (2018) testaram um algoritmo CNN profundo no diagnóstico e predição de doença periodontal aplicado a radiografias periapicais, obtendo resultados semelhantes àqueles observados com periodontistas experientes. A acurácia do CNN no diagnóstico de doença para pré-molares foi de 81% e, para molares, de 76,7%. Os maiores valores de acurácia foram observados em casos de comprometimento periodontal severo, e os valores mais baixos nos casos de comprometimento moderado. A acurácia de predição de exodontia em dentes clinicamente diagnosticados com doença periodontal severa foi de 82,8% para pré-molares e de 73,4% para molares. Foi observada acurácia diagnóstica maior do modelo para pré-molares do que para molares, o que poderia ser devido ao fato de os pré-molares apresentarem, geralmente, uma morfologia mais simples.

Em estudo testando CNN para detecção de perda óssea periodontal em radiografias panorâmicas, Kim *et al.* (2019) observaram que o modelo teve desempenho superior ao dos especialistas humanos, com F1 de 0.75 e AUC de 0.95 comparados a valores de 0.69 e 0.85 dos dentistas, respectivamente. Contudo, consideram que ainda é necessária maior validação clínica para adoção do método proposto na prática clínica.

Em trabalho semelhante, Krois *et al.* (2019) aplicaram CNN para diagnóstico de perda óssea periodontal em panorâmicas, e observaram que, quanto à acurácia, não houve diferença estatisticamente significativa entre o modelo testado e os especialistas. Ressalta-se, contudo, que este estudo apresenta algumas limitações, o que, sem dúvida, teve reflexos nos resultados obtidos. As principais limitações referem-se ao uso de uma amostra pequena associada a uma CNN de arquitetura superficial, com sete camadas; utilização de panorâmicas para diagnóstico periodontal e ausência de um grupo de teste independente, já que a amostra foi dividida em dados de treinamento (1.400 imagens) e de validação (350 imagens), apenas.

O estudo de Chang *et al.* (2020) se diferencia dos anteriores ao não se limitar à detecção das regiões com perda óssea em panorâmicas, realizando a quantificação e qualificação dessas perdas em cada dente individualmente e, assim, indicando o estagiamento da doença, segundo os novos critérios estabelecidos pelo *World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions*, de 2017. Este estagiamento, relacionado à severidade e extensão da periodontite, é determinado através da mensuração clínica da perda de inserção ou, caso esta não seja possível, através da detecção da perda óssea radiográfica. Contudo, a mensuração manual das perdas ósseas em todos os dentes em uma radiografia panorâmica se torna, frequentemente, uma tarefa trabalhosa e que pode consumir muito tempo do profissional. Assim, os autores propõem um sistema automatizado para avaliação das perdas ósseas periodontais e estagiamento da doença em radiografias panorâmicas, em cada dente individualmente. Os índices de acurácia diagnóstica para o modelo avaliado foram de 0.93 em relação à perda óssea periodontal, 0.91 em relação à junção cimento-esmalte e 0.91 para o dente analisado. Os resultados foram comparados aos índices de três radiologistas, calculando-se a correlação de coeficientes Pearson (PCC) e o coeficiente de correlação intraclass (ICC). Obteve-se, assim, um PCC global de 0.73, indicando forte correlação entre os índices dos radiologistas e os do modelo, e um ICC global de 0.91, demonstrando excelente confiabilidade do desempenho diagnóstico do modelo para perda óssea periodontal. Apesar dos bons resultados, os autores consideram que este método não fornece valores absolutos de perda óssea periodontal devido à magnificação e distorção de imagem nas radiografias panorâmicas, sugerindo futuros estudos com avaliação através de colaboração interorganizacional. Observa-se, ainda, outras limitações, como a amostra reduzida para o grupo teste e a inclusão da análise de imagens com baixa qualidade, como as de incisivos com sobreposição da imagem da coluna vertebral, o que se refletiu na acurácia mais baixa do modelo no diagnóstico para este grupo dentário.

A aplicação de CNN para detecção automática de dentes comprometidos periodontalmente em radiografias panorâmicas também foi testada por Thanathornwong e Suebnukarn (2020), obtendo índice médio de precisão de 0.81, índice médio de *recall* de 0.80, sensibilidade de 0.84, especificidade de 0.88 e índice F-1 de 0.81, o que indicaria um desempenho satisfatório do modelo. Este

desempenho, contudo, não pôde ser comparado diretamente ao desempenho dos três periodontistas de referência, pois os dentes por eles identificados como tendo comprometimento periodontal não foram diagnosticados apenas através de imagens radiográficas, mas também através de exame clínico periodontal.

Durante a análise dos trabalhos, não ficou clara a preferência da maioria deles por radiografias panorâmicas para diagnóstico de doença periodontal. Kim *et al.* (2019) chegam a afirmar que a detecção de perda óssea neste tipo de exame pode ser uma tarefa difícil, com baixos índices de concordância intra e interexaminadores devido à estrutura complexa e à baixa resolução mostradas nas imagens.

Apesar do bom desempenho geral dos modelos, alguns autores ressaltam que é impossível fazer um diagnóstico completo e previsão de doença periodontal usando apenas radiografias periapicais bidimensionais (LEE *et al.*, 2018; KROIS *et al.*, 2019), sendo necessário rever de forma abrangente dados radiográficos e clínicos, como histórico do paciente, profundidade de sondagem clínica, nível de inserção clínica, sangramento na sondagem, mobilidade, percussão e teste pulpar. Portanto, um algoritmo CNN profundo usando imagens radiográficas periapicais por si só não forneceria evidências suficientes, embora ainda possa ser usado como referência para o diagnóstico e previsão de doença.

A busca de trabalhos na área de Implantodontia resultou em seis artigos, todos do ano de 2020, e todos com foco na identificação de sistemas de implantes presentes em imagens radiográficas.

A correta identificação do tipo de sistema de fixação utilizado no implante se torna importante especialmente em situações de manutenção do trabalho realizado. Contudo, estes sistemas sofrem muitas modificações ou substituições ao longo do tempo, de forma que o clínico pode ter dificuldade em reconhecê-los, bem como suas intercompatibilidades, pela simples visualização nas radiografias, caso o prontuário do paciente não esteja disponível para consulta. Por essa razão, pode ser difícil encontrar a substituição mais adequada para um sistema mesmo quando ocorrem complicações comuns com os implantes, como afrouxamento ou fratura do parafuso (KIM *et al.*, 2020).

De fato, na ausência do prontuário, o conhecimento sobre o tipo de implante só seria possível através das radiografias, pois a maior parte do sistema de fixação do implante fica integrada ao osso alveolar, impedindo sua visualização no exame oral.

Assim, a identificação radiográfica dos implantes é especialmente importante para fornecer diagnósticos e tratamentos adequados aos pacientes. Lee *et al.* (2020) consideram que, se ocorrerem complicações biológicas ou mecânicas, a identificação do sistema utilizado é importante para minimizar intervenção invasiva. Por outro lado, distinguir radiograficamente características específicas do sistema de implante utilizado pode ser tarefa difícil, pois muitos apresentam similaridades, além de as panorâmicas e periapicais possuírem limitações inerentes.

Nesse sentido, Kim *et al.* (2020) testaram o desempenho de cinco modelos diferentes de CNN (SqueezeNet, GoogLeNet, ResNet-18, MobileNet-v2 e ResNet-50) na identificação de quatro sistemas diferentes de fixação de implantes (Branemark, Dentium Implantium, Straumann Bone Level e Straumann Tissue Level) em radiografias periapicais. Os parâmetros medidos, para fim de comparação entre os modelos, foram acurácia, precisão, revocação (*recall*) e score F1, com resultados de mais de 0.90 para todos eles, mostrando que as CNNs investigadas foram capazes de classificar quatro sistemas de implante com alta precisão, apesar da rede relativamente pequena e do pequeno número de imagens. Portanto, os sistemas analisados têm potencial para ajudar profissionais e pacientes a evitar tratamentos e despesas desnecessárias, resultantes do desconhecimento do tipo exato de implante utilizado.

Na fase de reconhecimento das características das imagens, foi utilizada a metodologia de mapeamento de ativação de classe (*class activation map* – CAM), que mostra, discriminada e claramente na imagem, a região usada pelo CNN para identificar cada sistema dentro de uma categoria. Assim, cada modelo buscou por uma parte distinta de cada implante analisado para fazer a classificação. De acordo com os CAMs da ResNet-50 e Squeezenet, que mostraram a melhor localização, as redes buscaram a conexão entre o sistema de fixação e o *abutment* para os implantes Brånemark, a área de fixação geral nos implantes Dentium e Straumann (BL), e a porção transgengival e conexão de fixação nos implantes Straumann (TL). Todas essas partes são distintas para cada tipo de implante. Esses achados comprovariam que o modelo de aprendizagem profunda pode identificar com precisão as características discriminatórias de cada tipo de implante.

Apesar de ser geralmente considerado que, quanto mais profunda a rede, mais profundo o aprendizado e melhor o desempenho do modelo, os autores ressaltam que

essa tendência não ficou evidente nos resultados deste experimento. O aumento do número de camadas de rede não foi necessariamente proporcional ao aumento da precisão, podendo-se considerar que o aprendizado foi suficiente mesmo com menor número de camadas. Os autores consideram, ainda, que este estudo confirmou que uma CNN foi capaz de analisar imagens de implante com precisão e classificar automaticamente os quatro tipos de implante selecionados, mesmo com um tamanho de rede relativamente pequeno e um pequeno número de imagens. Dessa forma, as redes de classificação de implantes poderiam ser facilmente aprendidas com poder computacional relativamente baixo e serem aplicadas em diferentes situações, tornando-as úteis e convenientes em situações clínicas.

Os autores apontam, ainda, algumas limitações do estudo e melhorias a serem postas em prática em trabalhos futuros, como a inclusão e análise de maior diversidade de sistemas de implantes e a criação de uma rede capaz de: detectar sistemas de implante usando imagens não segmentadas, ou aplicar uma técnica para detectar vários implantes simultaneamente; aprender e identificar tipos de implantes usando radiografias panorâmicas e classificar não apenas os tipos de implante, mas também seus diâmetros e comprimentos. Se o comprimento do implante puder ser detectado automaticamente, o grau de perda óssea marginal ao redor do implante pode ser facilmente verificado, o que, por sua vez, pode levar ao desenvolvimento de um algoritmo para estimar a higidez e o prognóstico do implante, bem como diagnosticar peri-implantite, levando ao desenvolvimento de um software de diagnóstico clinicamente útil para complicações relacionadas ao implante. Além disso, ao desenvolver uma rede que classifique com precisão o diâmetro do implante, o sistema de implante poderia ser automaticamente identificado, e seria possível saber quais componentes devem ser preparados para reparo e manutenção quando ocorressem complicações mecânicas.

Lee e Jeong (2020) realizaram estudo no qual testaram um modelo CNN (GoogLeNet Inception-v3) quanto à capacidade de identificação e classificação de três sistemas de implantes com características similares (TSIII AS, Superline e Straumann BL), em radiografias panorâmicas e periapicais, e compararam seu desempenho com o de um especialista periodontista. Os resultados mostraram que o modelo teve desempenho altamente preciso na identificação dos implantes, com AUC entre 0.956 e 0.979, tendo superado o desempenho do periodontista (AUC entre 0.891 e 0.959).

Observou-se que o modelo mostrou maior acurácia na identificação do sistema Straumann BL, o que os autores acreditam dever-se à maior conicidade desse sistema. As limitações apontadas pelos autores acerca do estudo foram a amostra relativamente pequena, sendo necessária a técnica de *data augmentation*; o número pequeno de sistemas de implante para identificação (apenas três tipos de sistema de implante) e, finalmente, a utilização de imagens em duas dimensões para análise pelo modelo.

Um modelo de CNN para identificação de modelos de implantes em radiografias periapicais e panorâmicas foi desenvolvido no estudo de Said *et al.* (2020), com acurácia diagnóstica de 93,8%. Os resultados foram superiores para implantes com características discriminatórias. Os autores ressaltam que os *designs* dos implantes incluídos no estudo foram bem diferentes entre si, o que poderia explicar os excelentes resultados obtidos.

O estudo de Lee *et al.* (2020) também teve como foco a classificação de sistemas de implantes em radiografias panorâmicas e periapicais utilizando um modelo CNN, cujo desempenho foi comparado com o de vinte e cinco profissionais. Foram utilizadas 11.980 imagens contendo seis diferentes tipos de sistemas de implantes: Astra OsseoSpeed TX® (Dentsply), Implantium® (Dentium), Superline® (Dentium), TSIII® (Osstem), SLActive BL® (Institut Straumann AG), SLActive BLT® (Institut Straumann AG). A CNN utilizada foi do tipo profunda automatizada, capaz de selecionar automaticamente o modelo de aprendizado profundo e fazer a otimização automática dos hiperparâmetros a serem aplicados. Os resultados apontaram um excelente desempenho geral do modelo, com índices globais AUC de 0.954, sensibilidade de 0.955 e especificidade de 0.853, superando o desempenho dos especialistas. Não houve diferença estatisticamente significativa na análise quanto ao tipo de radiografia (panorâmicas, periapicais e panorâmicas somadas às periapicais). Quanto aos tipos de sistemas de implante analisados, o sistema de implante Straumann SLActive® BLT, por ter uma forma relativamente grande em comparação com os outros, obteve classificações adequadas com AUC elevado por parte do modelo e dos profissionais. No entanto, o Dentium Superline® e o Osstem TSIII®, sistemas de implante que não têm elementos característicos evidentes com uma forma afilada, foram identificados adequadamente pela CNN automatizada, enquanto os profissionais odontológicos apresentaram baixo AUC. Com base nesses

resultados, a CNN mostrou precisão de classificação significativamente maior do que os profissionais, incluindo periodontistas experientes, tendo mostrado alta eficácia na classificação de sistemas de implante semelhantes com base em imagens radiográficas. Além disso, o trabalho mostrou que a diferença no nível de experiência dos profissionais associado aos sistemas de implante não afetou significativamente a precisão de classificação destes. Os autores apontam como limitações do estudo um insuficiente conjunto de dados para treinamento do modelo, sendo necessária a construção de um conjunto de dados de alta qualidade e larga escala contendo diferentes tipos de sistemas de implante, e a utilização de imagens bidimensionais, sugerindo, para estudos posteriores, a adoção de imagens tridimensionais, o que poderia resultar em uma melhora significativa da acurácia do modelo.

Sukegawa *et al.* (2020) analisaram a acurácia diagnóstica de cinco modelos de CNN profunda na identificação de onze tipos de implantes em radiografias panorâmicas, avaliando a proporção de classificações corretas. Para detectar potenciais classificações corretas com base em características incorretas e obter alguma intuição no processo de classificação, foram identificados os pixels de imagem mais relevantes para classificação usando mapas de ativação de classe ponderados por gradiente, metodologia também utilizada no estudo de Kim *et al.* (2020). Nos mapas, as cores "mais quentes" representam as regiões de maior importância utilizadas pelo modelo para a classificação, ajudando a entender que características ou áreas de imagens são usadas para decisões de classificação. Os resultados mostraram que todos os modelos apresentaram alta acurácia diagnóstica, com índices entre 0.860 e 0.935, mostrando-se altamente eficazes na identificação dos sistemas de implantes. Os autores sugerem que futuramente, visando a aplicabilidade clínica destes modelos, será necessário estimar os resultados em um grande estudo transversal envolvendo várias radiografias panorâmicas com diferentes qualidades de imagem, ou mesmo com diferentes tipos de implantes simultaneamente.

De forma geral, a análise dos resultados dos estudos sobre aplicação de CNNs à identificação dos tipos de sistemas de implantes em radiografias periapicais e panorâmicas confirmou a efetividade dos modelos. Quando voltada para este objetivo, a aplicação de modelos de inteligência artificial, por si só, mostrou potencial como valioso auxiliar do clínico, sendo capaz de identificar diferenças mesmo entre sistemas

de configuração semelhantes, com resultados similares tanto em panorâmicas quanto em periapicais.

Além dos estudos apresentados até aqui, a busca bibliográfica realizada também apontou resultados da aplicação de modelos de IA na análise de imagens radiográficas com outros propósitos. Um deles, bastante relevante para a rotina de trabalho do radiologista oral, seria o diagnóstico de lesões císticas e tumorais em mandíbula e maxila.

Uma ferramenta auxiliar precisa, rápida e confiável para o diagnóstico automático nesta área seria de alto valor na prática clínica, já que diversas lesões apresentam características radiográficas similares, mas comportamentos muito distintos, o que impacta diretamente no prognóstico e na tomada de decisão quanto ao tratamento. Além disso, o diagnóstico das lesões, se feito em seus estágios iniciais, pode evitar cirurgias mais extensas e complexas.

Dessa forma, alguns trabalhos (Anexo VIII, Tabela VIII) tiveram como objetivo testar o desempenho de redes neurais na detecção e classificação de cistos e tumores odontogênicos em radiografias de maxila e mandíbula.

Poedjiastoeti e Suebnukarn (2018) criaram e testaram uma CNN para detecção de ameloblastomas e ceratocistos odontogênicos em panorâmicas, comparando os resultados com aqueles de cinco especialistas bucomaxilofaciais. A habilidade da CNN no diagnóstico destas lesões foi similar à dos especialistas, ambos com acurácia de 83%. O diferencial foi o tempo requerido para o diagnóstico, muito menor para a rede neural, que precisou de 38 segundos, enquanto os especialistas precisaram de 23 minutos, em média.

Ariji *et al.* (2019) testaram CNN na identificação e na classificação de lesões radiolúcidas mandibulares em panorâmicas. No estudo, foram incluídas cinco lesões de ocorrência frequente em mandíbula: ameloblastomas, ceratocistos odontogênicos, cistos dentígeros, cistos radiculares e cistos ósseos simples. Os critérios de inclusão foram a presença de lesão radiolúcida na mandíbula e verificação histopatológica do diagnóstico. Todas as lesões tinham pelo menos 10 mm de diâmetro. Os resultados mostraram que o modelo foi capaz de detectar corretamente a presença de lesão em 88% das imagens. Os autores utilizaram o termo sensibilidade de detecção para referirem-se à taxa em que o modelo de aprendizagem detectou corretamente a lesão. O termo sensibilidade de classificação referiu-se à taxa em que o modelo classificou

corretamente a lesão. A sensibilidade de detecção e sensibilidade de classificação foram de 0,71 e 0,60, respectivamente, para os ameloblastomas; 1,00 e 0,13 para ceratocistos odontogênicos; 0,88 e 0,82 para cistos dentígeros, e 0,81 e 0,77 para cistos radiculares. Todos os ceratocistos odontogênicos foram detectados, mas a sensibilidade de classificação foi muito baixa. O melhor desempenho para detecção e classificação foi para cistos dentígeros, o que os autores creditam à sua localização preferencial. Para a detecção das lesões, o modelo testado analisou principalmente as informações sobre os limites da lesão e sua posição na estrutura anatômica. As lesões melhor detectadas foram aquelas de limites mais bem definidos.

Nos casos de falha de detecção, as lesões não puderam ser detectadas devido ao seu pequeno tamanho, bordas mal definidas e/ou imagem levemente radiolúcida. Pequenas lesões fornecem menos informações para a discriminação de suas classes e, portanto, mais dados de treinamento seriam necessários para detectar com sucesso tais lesões. Os autores consideram que, em casos de pouco tempo disponível para o diagnóstico por imagem ou, ainda, de pouca experiência por parte do profissional, a detecção automática de lesões é considerada clinicamente mais útil do que a classificação de imagem. De fato, uma detecção eficiente poderia contribuir para reduzir falhas de diagnóstico por não visualização de imagens não relacionadas à queixa principal do paciente.

Uma limitação do estudo apontada pelos autores diz respeito à necessidade de identificar manualmente cada lesão na fase de treinamento do modelo, o que se torna demorado e trabalhoso. Este processo foi realizado por apenas um radiologista especialista e, portanto, não foi possível avaliar a variabilidade interobservador. No futuro, a segmentação automatizada será necessária, e se ela se provar rápida, reproduzível e fácil de usar, será aplicada na prática clínica.

Além de demonstrar alto desempenho diagnóstico para lesões radiolúcidas em mandíbula analisadas em panorâmicas, o modelo realizou os diagnósticos em treze segundos, em média, sugerindo potencial para tornar mais dinâmico e ágil o processo de diagnóstico neste tipo de exame.

Lee *et al.* (2020) incluíram imagens de tomografias computadorizadas de feixe cônico, além de panorâmicas, para testar a capacidade de uma rede CNN em detectar e diagnosticar três tipos de lesões radiolúcidas: ceratocistos odontogênicos, cistos dentígeros e cistos periapicais. Este estudo demonstrou que as lesões foram

efetivamente detectadas e diagnosticadas com base na arquitetura profunda da CNN tanto em panorâmicas quanto em TCFC, sendo que, nestas últimas, o desempenho diagnóstico foi maior do que quando treinado com imagens panorâmicas. Ao utilizar imagens panorâmicas, a precisão diagnóstica total foi de 84,6%; a precisão diagnóstica foi maior para cistos periapicais (87,0%) e menor para ceratocistos (81,8%). Para as TCFC, a acurácia total de diagnósticos foi de 91,4%; a precisão diagnóstica foi maior para cistos periapicais (93,7%) e menor para ceratocistos (87,2%). Segundo os autores, estes resultados parecem indicar que as imagens das tomografias têm vantagens importantes, como mais detalhes e menos artefatos nos limites anatômicos das áreas de interesse do que as imagens panorâmicas convencionais.

Os autores fazem a ressalva de que, apesar do bom desempenho da CNN, a acurácia do diagnóstico de lesões císticas odontogênicas através de imagens radiológicas, por si só, fica aquém daquela obtida com o exame histopatológico e permanece um desafio. No entanto, pode-se argumentar que o uso da tecnologia das redes neurais para fins diagnósticos não parece pretender substituir outras metodologias consagradas, mas sim complementá-las, e, do ponto de vista do radiologista que depende apenas da imagem radiográfica para lançar hipóteses diagnósticas, um sistema automatizado ágil e preciso seria um valioso aliado.

Kwon *et al.* (2020) realizaram pesquisa com o objetivo de diagnosticar automaticamente cistos odontogênicos e tumores de mandíbula e maxila em radiografias panorâmicas usando aprendizado profundo. As lesões incluídas foram cistos dentígeros, cistos periapicais, ceratocistos odontogênicos e ameloblastomas, com confirmação diagnóstica por histopatológico e pré-diagnosticadas por dois radiologistas experientes. Foram incluídas 100 imagens de mandíbula ou maxila sem doença, como controle. O desempenho global do modelo revelou acurácia de 95,6% para o diagnóstico de doença. O melhor desempenho diagnóstico foi para cistos dentígeros, com acurácia de 97,8%. A detecção de normalidade teve acurácia de 96%. Os autores observaram que a sensibilidade para lesões localizadas na maxila foi menor do que para lesões na mandíbula devido à sobreposição entre lesões e outras estruturas anatômicas na maxila.

O modelo de rede neural testado neste estudo mostrou, de forma geral, maior sensibilidade para as lesões analisadas do que o modelo de Arijji *et al.* (2019),

provavelmente porque esses autores usaram um número menor de imagens de lesões para treinamento e avaliação e, além disso, foram usadas imagens somente de mandíbula.

Em estudo semelhante ao de Kwon *et al.* (2020), Yang *et al.* (2020) testaram uma versão anterior de CNN para diagnóstico de cistos dentígeros, ceratocistos odontogênicos e ameloblastomas em panorâmicas. Na amostra de teste, foram incluídas imagens sem alterações patológicas, classificadas como “sem lesão”. O referencial comparativo seria o diagnóstico previamente confirmado por histopatológico e o diagnóstico feito por bucomaxilofaciais e clínicos gerais. O tempo médio para avaliação das imagens foi de 33,8 minutos para os humanos, enquanto para o modelo foi descrita detecção “em tempo real”, ou seja, instantânea. A acurácia diagnóstica do modelo foi de 66%, enquanto a dos especialistas variou de 64% a 66% e a dos clínicos gerais, ficou entre 58% e 60%. Apesar dos índices de precisão e *recall* do modelo terem sido mais altos que os dos humanos, a diferença se mostrou estatisticamente insignificante, com desempenho diagnóstico equivalente entre eles. Os autores observaram que o modelo demonstrou alto entendimento contextual das imagens, semelhante ao do sistema cognitivo humano, o que contribuiu para redução dos índices de falsos positivos. Por outro lado, sob determinadas circunstâncias, o modelo pode apresentar dificuldade no reconhecimento de imagens, já que cistos e tumores odontogênicos aparecem na radiografia panorâmica com várias características no corpo da lesão e nas bordas. Se essas configurações da patologia não forem incluídas na fase de treinamento do modelo, este pode não as reconhecer. Por exemplo, lesões extensas com corpos grandes e ocos apresentarão uma grande área de radiolucência em uma radiografia panorâmica. Como resultado, quanto maior a lesão, maior a probabilidade de que o modelo classifique a área como livre de lesão. Na maxila, algumas lesões aparecem obscurecidas devido à baixa densidade óssea e às várias estruturas anatômicas adjacentes que se sobrepõe à área da lesão. Nesses casos, tanto o modelo quanto dois terços dos profissionais falharam em detectar as lesões.

Dos estudos analisados, observa-se a utilização, em todos eles, de radiografias panorâmicas para diagnóstico das lesões. De fato, lesões císticas e mesmo tumorais são frequentemente identificadas como achados incidentais neste tipo de exame, muitas vezes sem sinais ou sintomas indicativos de sua presença, dependendo de

seu estágio de desenvolvimento. Um dos estudos (LEE *et al.*, 2020) utilizou também as tomografias computadorizadas de feixe cônico, obtendo melhor acurácia diagnóstica do modelo quando comparado aos resultados das panorâmicas. Contudo, as panorâmicas ainda constituem o exame de rotina preferencial, provavelmente por ser mais acessível aos pacientes e pelas menores doses de radiação empregadas.

Além disso, parece haver um consenso de que a aplicação das redes neurais para o diagnóstico das lesões poderia representar um suporte na prática clínica, no sentido de formular uma “segunda opinião” ou mesmo de reconfirmar a detecção e classificação dos cistos e tumores odontogênicos. A velocidade de detecção das lesões representaria uma vantagem considerável, mesmo que o desempenho diagnóstico tenha se mostrado similar ao dos humanos.

Em uma revisão sistemática da literatura e meta-análise por Silva *et al.* (2020), a avaliação geral mostrou que todos os estudos apresentaram altas taxas de precisão dos sistemas de diagnóstico na classificação de lesões maxilofaciais radiolúcidas em comparação com a biópsia histopatológica. No entanto, devido à heterogeneidade moderada encontrada entre os estudos incluídos na meta-análise, os autores consideraram que não foi possível uma recomendação pragmática sobre o uso de análise assistida por computador.

A aplicação de redes neurais convolucionais também foi encontrada em três trabalhos que avaliaram a acurácia diagnóstica de modelos na detecção de osteoporose em panorâmicas (LEE *et al.*, 2019; LEE *et al.*, 2020; NAKAMOTO *et al.*, 2020), através da análise de alterações na textura e características morfológicas do tecido ósseo. Segundo os autores, a utilização desta modalidade de exame para este fim apresenta boa relação custo/benefício e eficácia, sendo suficiente para embasar o encaminhamento do paciente para tratamento médico apropriado. Contudo, o diagnóstico presuntivo de osteoporose através de panorâmicas pode não ser tarefa fácil, especialmente para o clínico geral, pois se baseia na interpretação subjetiva das imagens. Portanto, o desenvolvimento de um sistema automatizado efetivo que pudesse auxiliar os profissionais nesta tarefa poderia permitir aos pacientes a implementação do tratamento inclusive nos estágios iniciais da doença, já que sinais e sintomas costumam estar ausentes até que haja uma maior deterioração do tecido ósseo, com risco aumentado de fratura. Segundo os autores, estudos anteriores confirmaram que a investigação da porosidade do osso cortical em panorâmicas seria

uma maneira efetiva de fazer a triagem de casos de osteoporose e o prognóstico de ocorrência de fratura vertebral.

Lee *et al.* (2019) testaram uma rede neural convolucional em radiografias panorâmicas para detecção de osteoporose e compararam os resultados com os de radiologistas experientes. A região analisada para o diagnóstico foi a região central da base mandibular dos lados esquerdo e direito. Os resultados mostraram que os valores de acurácia variaram de 92,5% a 98,5%, tendo sido observada alta concordância entre o modelo e os especialistas.

Em outro estudo, Lee *et al.* (2020), usando a técnica de mapeamento de ativação de classe ponderada por gradiente, uma interpretação visual do modelo CNN profundo de melhor desempenho indicou que o modelo se baseava em características de imagem na borda inferior esquerda e direita da mandíbula. De acordo com essa interpretação visual, os índices de acurácia diagnóstica foram elevados quando as características da imagem foram especificadas na região média do lado esquerdo e direito da borda inferior mandibular. Essa região também é consistente com as regiões utilizadas para discriminar a osteoporose usando imagens de panorâmicas, na maioria dos estudos anteriores, segundo os autores. Assim, confirmou-se a região central da base mandibular dos lados esquerdo e direito como preferenciais para análise do tecido ósseo para osteoporose. Os autores consideram que os resultados sugerem que a avaliação baseada em aprendizagem profunda das imagens de panorâmicas pode ser útil e confiável na triagem automatizada de pacientes com osteoporose.

O estudo de Nakamoto *et al.* (2020) desenvolveu e testou um sistema de diagnóstico auxiliado por computador (*computer-aided diagnosis system – CAD*) para detectar osteoporose em imagens de panorâmicas, obtendo acurácia de 75% quando comparado com exames de densidade óssea da coluna lombar. O novo sistema utilizou o índice de severidade de reabsorção óssea linear (BRSI) para avaliar quantitativamente a porosidade do osso cortical marginal inferior mandibular. A avaliação visual da porosidade na borda inferior da mandíbula e do BRSI tiveram alta correlação. Assim, concluiu-se que, quanto maior o BRSI, maior a porosidade mandibular. Foi encontrada forte correlação negativa entre o BRSI e a densidade óssea da coluna lombar. Assim, considerou-se que, quanto maior o BRSI, maior a possibilidade de baixa densidade óssea vertebral. Os autores consideram que o novo sistema de triagem é capaz de avaliar quantitativamente a porosidade cortical

mandibular, permitindo o rastreamento preventivo da osteoporose e aumentando, assim, as perspectivas clínicas.

Outra aplicação clínica encontrada para as redes neurais foi no diagnóstico de sinusite maxilar, com dois estudos selecionados. Em um deles (Kim *et al.*, 2019), os autores testaram um algoritmo de aprendizado profundo na detecção de sinusite maxilar em radiografias de Waters, comparando os resultados com o diagnóstico feito por radiologistas. Os autores consideram que o algoritmo foi capaz de diagnosticar sinusite maxilar com AUC superior e sensibilidade e especificidade comparáveis às dos radiologistas. O outro estudo (MURATA *et al.*, 2019), com metodologia semelhante, mas utilizando panorâmicas, também não mostrou diferença estatisticamente significativa entre o desempenho diagnóstico do modelo e aquele dos radiologistas. Ambos consideraram a habilidade diagnóstica de seus modelos suficientemente alta para prover apoio diagnóstico na clínica, especialmente para profissionais inexperientes.

Quanto à modalidade dos exames utilizados, Kim *et al.* (2019) afirmam que a TCFC é a modalidade de escolha para diagnóstico de sinusite, por fornecer melhor detalhamento anatômico dos seios paranasais. Contudo, a radiografia de Waters foi utilizada por mostrar razoável acurácia no diagnóstico de sinusite no seio maxilar, ao contrário de sinusite em outros locais, e ser mais acessível. A TCFC foi utilizada como referencial padrão. Murata *et al.* (2019) utilizaram panorâmicas por considerarem esta modalidade como a mais utilizada para triagem odontológica. As limitações do exame, como sobreposição de estruturas anatômicas na região do seio maxilar, podem tornar difícil o diagnóstico de sinusite, principalmente para profissionais inexperientes, o que justificaria a utilização de um sistema automatizado como auxiliar de diagnóstico.

No estudo de Kim *et al.* (2019), todas as radiografias foram rotuladas de acordo com 6 categorias: 0, normal; 1, espessamento da mucosa do seio maior que 4 mm; 2, presença de nível aéreo; 3, opacificação total; 4, casos com complicação (por exemplo, cisto de retenção); e 5, incapaz de ser rotulado (incluindo casos de atelectasia do seio). Quando um paciente exibia características de mais de 1 rótulo, o rótulo mais alto foi atribuído. Um cisto de retenção (rótulo 4) era considerado quando uma lesão com radiopacidade semelhante a tecido mole, de formato arredondado ou em forma de cúpula estava presente, especialmente no soalho do seio maxilar, sem nível de fluido de ar ou opacificação total. A atelectasia sinusal (rótulo 5) foi

considerada quando o diâmetro do seio maxilar era menor que 8 mm. As mudanças escleróticas nas paredes ósseas do seio maxilar por si só não foram consideradas como sinusite. Não só o rótulo 5, mas também o rótulo 4 foi excluído nas análises posteriores, pois muitas vezes era impossível determinar a presença de sinusite coexistente sobreposta a um cisto de retenção. Os resultados mostram que tanto o modelo quanto os radiologistas tiveram dificuldade em diferenciar casos de Rótulo 1 dos de Rótulo 0. Os autores ressaltam que sinusite não poderia ser diagnosticada apenas através de imagem, porque o espessamento mucoso dos seios maxilares também aparece na presença de infecção do trato respiratório superior menor.

O emprego das redes neurais convolucionais também se estendeu à tarefa de identificação e numeração de elementos dentários em radiografias. Esta se configura em tarefa rotineira para o radiologista, podendo ocupar várias horas valiosas e causar erros de diagnóstico devido a fatores pessoais, como fadiga e baixos níveis de experiência. Tanto a carga de trabalho quanto os erros de diagnóstico poderiam ser reduzidos se ferramentas inteligentes de interpretação de radiografias fossem desenvolvidas para melhorar a qualidade da assistência odontológica. Nessa perspectiva, a identificação automática de dentes usando radiografias digitais se configura em valiosa ferramenta auxiliar para o radiologista.

Zhang *et al.* (2018) testaram o desempenho de CNN com este fim em radiografias periapicais, obtendo precisão do modelo de 95,8%. Os autores afirmam que seu método pode lidar com muitos casos complexos, como áreas com perdas dentárias, elementos com prótese fixa, com lesões de cárie e com implantes, o que não aconteceria com sistemas de numeração tradicionais, baseados no contorno e localização dos dentes.

Chen *et al.* (2019) também testaram CNN em radiografias periapicais para detecção e numeração de dentes, comparando os resultados com aqueles de três dentistas, obtendo índices de precisão e de *recall* superiores a 90%. Os autores consideram que o desempenho diagnóstico do modelo foi similar ao de um dentista recém-formado.

Tuzoff *et al.* (2019) propuseram um novo sistema baseado em CNN para detecção e numeração de dentes em panorâmicas, obtendo índices de precisão e sensibilidade entre 98% e 99%, equivalentes aos dos especialistas. Os autores consideram que, com base nesses achados, o método tem potencial para aplicação

prática e posterior avaliação para análise automatizada de radiografias dentárias, simplificando o processo de preenchimento de prontuários digitais e ajudando a economizar tempo.

Ainda na área de detecção e numeração de dentes, o estudo de Kuwada *et al.* (2020) buscou utilizar modelos de IA para identificar e classificar dentes supranumerários impactados na região anterior da maxila em pacientes com incisivos totalmente erupcionados. Os autores consideram que imagens na região de incisivos superiores são muitas vezes difíceis de interpretar, devido à sobreposição da coluna cervical e à possíveis distorções nesta região, o que pode prejudicar a detecção de supranumerários. Foram testados três modelos, DetectNet, VGG-16 e AlexNet. O DetectNet produziu os mais altos valores de eficácia diagnóstica. O VGG-16 rendeu valores significativamente mais baixos em comparação com o DetectNet e o AlexNet. A avaliação do desempenho de detecção do DetectNet mostrou que o *recall*, a precisão e o índice F1 foram todos 1.0, indicando detecção perfeita. Os autores consideram que os resultados indicaram potencial dos sistemas de aprendizagem profunda para fornecer suporte diagnóstico na interpretação das radiografias panorâmicas. Contudo, apontam como limitação, além da amostra de treinamento relativamente pequena, problema comum a quase todos os estudos desta revisão, a não inclusão de pacientes com dentição mista. A detecção de supranumerários neste tipo de dentição seria mais difícil devido à presença de dentes permanentes não irrompidos.

Observa-se que os estudos sobre detecção e classificação automática de elementos dentários em radiografias, tal como visto em estudos de aplicação de IA em outras áreas, buscam agilizar o processo na prática clínica, ao mesmo tempo em que forneceriam resultados confiáveis que serviriam como confirmação do diagnóstico ou da análise realizada pelo profissional. Nesse sentido, um dos estudos (TUZOFF *et al.*, 2019) mostrou resultados de desempenho do modelo compatíveis com potencial para futura utilização na rotina clínica.

Outra aplicação das redes neurais no diagnóstico radiográfico focou na detecção automática de terceiros molares e canal mandibular, e na relação entre essas estruturas, com o objetivo de facilitar a avaliação do risco de dano ao nervo alveolar inferior quando da remoção dos terceiros molares e proporcionar maior segurança quando da fixação de implantes.

Nesse sentido, Vinayahalingam *et al.* (2019) desenvolveram e testaram uma CNN quanto à habilidade em detectar terceiros molares e canal mandibular em radiografias panorâmicas, comparando os resultados com aqueles da segmentação manual. O coeficiente Dice do modelo para detecção de terceiros molares inferiores teve média de 93,6%, e para detecção do canal mandibular, de 80,5%. Foi realizada, adicionalmente, inspeção visual das segmentações manual, automatizada e uma sobreposição das duas. Essa inspeção, somada aos índices de sensibilidade e de especificidade obtidos, permitiu aos autores considerarem o desempenho do modelo como satisfatório. No entanto, em alguns casos, a CNN não conseguiu localizar e segmentar ambas as estruturas anatômicas predeterminadas de forma satisfatória. Os autores consideram que vários fatores estariam associados ao baixo desempenho, como a falta de contraste, nas panorâmicas, entre a mandíbula e o canal mandibular, complicando a tarefa de segmentação tanto para o observador quanto para a CNN, e a variabilidade da forma e localização do canal mandibular entre os pacientes. Assim, seriam necessários mais estudos comparativos, a fim de avaliar a qualidade dos métodos de segmentação antes da implementação clínica. Para os autores, os resultados encorajadores obtidos a partir do presente estudo na segmentação dos terceiros molares e do canal mandibular seriam o primeiro passo no caminho para implementar com sucesso o *deep-learning* na prática clínica diária.

Fukuda *et al.* (2020) testaram três redes neurais com diferentes números de camadas quanto à habilidade de análise da relação entre terceiros molares e canal mandibular em panorâmicas, que foram recortadas para mostrar a região de interesse (RDI) a ser analisada. Estes recortes foram feitos em dois tamanhos diferentes, a fim de determinar-se o efeito do tamanho dos recortes de treinamento no desempenho dos modelos. O estudo se diferencia também por comparar o tamanho de armazenamento do modelo treinado e a exigência de tempo do processo de treinamento entre as 3 CNNs, cujos resultados foram: para o AlexNet, 63,8GB e 30 min; para o GoogLeNet, 11,6GB e 2h; e, para o VGG-16, 62,4GB e 1h. Foi observado que o tempo e o espaço de armazenamento necessários para o processamento de aprendizagem dependeram de forma diretamente proporcional da profundidade das camadas das redes neurais e do número total de parâmetros aprendidos no processo, respectivamente. Quanto ao desempenho diagnóstico relacionado ao tamanho dos recortes, de forma geral, os modelos tiveram melhor desempenho nos recortes de

imagem menores, com valores de AUC variando de 0.88 a 0.93, enquanto os valores para os recortes maiores variaram de 0.75 a 0.93. Os autores acreditam que isso se deve provavelmente à quantidade de informação desnecessária incluída nos recortes de tamanho maior, sugerindo que deva ser dada maior atenção ao tamanho dos recortes das imagens quando da criação de um modelo de treinamento. Dos modelos testados, O GoogLeNet aplicado aos recortes menores teve o melhor valor da AUC, de 0.93. Concluiu-se que a técnica de aprendizagem profunda parece ter potencial para classificar a relação entre o terceiro molar inferior e o canal mandibular em radiografias panorâmicas, e pode ser útil para minimizar o risco de danos ao nervo alveolar inferior durante exodontias.

Jaskari *et al.* (2020) realizaram estudo utilizando CNN para localização automática do canal mandibular em imagens de TCFC, visando a inserção de implantes na mandíbula. A localização precisa de canais mandibulares e sua relação com os implantes que serão fixados é essencial para evitar danificar o nervo alveolar inferior, e esta localização é atualmente determinada manualmente a partir de imagens de TCFC, numa tarefa laboriosa e demorada. No estudo, foram utilizadas imagens 3D para determinação da localização do canal mandibular. A precisão da previsão do modelo foi da ordem de 0,5 mm para cerca de 90% do comprimento do canal mandibular, o que os autores consideram mais do que suficiente para implantodontia. Houve diminuição do desempenho na região próxima ao forame mandibular, região essa irrelevante do ponto de vista da implantodontia. Os autores concluem que um sistema automatizado baseado em rede neural de aprendizagem profunda, quando aplicado às varreduras de TCFC, pode produzir segmentações de alta qualidade de canais mandibulares.

Kwak *et al.* (2020) testaram três CNN na detecção do canal mandibular em imagens 2D e 3D de TCFC, obtendo índices de acurácia entre 0.76 e 0.99, com os maiores índices para imagens 3D. A rede aplicada em 3D, apesar dos melhores resultados, necessitou de maior poder computacional e maior tempo para a fase de treinamento. Os autores consideram este como um estudo preliminar capaz de encorajar novas oportunidades de aplicar o aprendizado profundo na segmentação de estruturas pequenas e complexas utilizando imagens de TCFC.

As redes neurais convolucionais também foram testadas na detecção de placas ateroscleróticas nas carótidas em radiografias panorâmicas, com o intuito de auxiliar

na prevenção de acidente vascular cerebral (KATS *et al.*, 2019). Os resultados mostraram sensibilidade de 75%, especificidade de 80% e acurácia de 83%. Segundo os autores, o fato de a imagem radiográfica dos ateromas poder demonstrar uma grande variedade de contornos e tamanhos e poder ser constituída de várias pequenas radiopacidades pode prejudicar severamente o desempenho do modelo, que precisa de uma delimitação clara de contornos. Outro fator a ser considerado seria o diagnóstico diferencial de estruturas anatômicas calcificadas na região, como a cartilagem tritícea calcificada ou calcificação do corno superior da cartilagem tireóidea. Os autores consideram os resultados promissores e sugestivos de potencial para futura utilização na prática clínica, necessitando estudos adicionais para aperfeiçoamentos do modelo.

Fukuda *et al.* (2020) testaram CNN na detecção de fratura radicular vertical em panorâmicas, abrangendo todos os grupos dentários, e obtiveram *recall* de 0.75, precisão de 0.93 e score F1 de 0.83. Foi observado que a detecção das fraturas foi maior no grupo dos molares inferiores e nos dentes com tratamento endodôntico. Os autores consideram que o desempenho do modelo pareceu insuficiente para aplicação na prática clínica. Uma limitação importante deste estudo se refere à inclusão somente de imagens radiográficas com linhas de fratura bem claras e distinguíveis nas raízes, situação que evidentemente não reflete a realidade da prática do radiologista e que tornaria dispensável um sistema auxiliar de diagnóstico.

Dois estudos testaram modelos de IA quanto ao desempenho na detecção de osteoartrite em articulação temporomandibular (ATM). O diagnóstico preciso desta desordem pode ser desafiador durante os estágios iniciais da doença e, como sinais e sintomas têm fraca correlação com osteoartrite de ATM, o exame clínico nem sempre parece ser suficiente para o diagnóstico.

Lee *et al.* (2020) utilizaram imagens de TCFC devido a esta modalidade de exame ser considerada o padrão-ouro para o diagnóstico de alterações ósseas da articulação temporomandibular, já que a integridade do osso cortical condilar e da estrutura óssea subjacente podem ser claramente observadas, além de fornecer dose de radiação menor, mas com precisão diagnóstica semelhante à tomografia convencional. O objetivo do estudo foi desenvolver uma ferramenta diagnóstica utilizando IA para detecção de reabsorção condilar em imagens de TCFC. Foram utilizados cortes sagitais de ATM, com recorte da região de cabeça de mandíbula,

cujas imagens apresentaram sinais de osteoartrite, como contorno irregular da cabeça da mandíbula, defeitos ósseos, perda de corticalização e achatamento da cabeça. Imagens que mostravam um contorno liso e arredondado do côndilo sem defeitos subcondrais foram consideradas normais e excluídas. As imagens, analisadas por um ortodontista experiente, foram classificadas em três categorias: classe 3, sem osteoartrite; classe 1, indeterminada para osteoartrite e classe 2, com osteoartrite. As imagens classificadas como classe 1 podiam corresponder a variação anatômica normal, variação relacionada com a idade do paciente, ou remodelação fisiológica. O desempenho do modelo mostrou índices médios de acurácia de 0.86, precisão de 0.85, *recall* de 0.84 e *score F1* de 0.84. Observa-se, neste estudo, que a utilização das imagens de TFCF se limitou aos cortes sagitais; logo, as imagens analisadas foram em duas dimensões, e não em três, como pode sugerir a modalidade de exame escolhida. No entanto, os autores consideram que foi possível observar a integridade do revestimento ósseo cortical e existência de esclerose subcondilar, que podem estar associadas ao estágio inicial da osteoartrite. O fato de alguns cortes sagitais poderem não mostrar alterações ósseas sugere a necessidade de observar também os cortes coronal e axial para um diagnóstico preciso.

A análise geral dos estudos incluídos nesta revisão permite inferir que um importante ponto a ser levado em conta nas pesquisas com IA diz respeito à aquisição dos conjuntos de dados necessários para treinamento e teste dos modelos. Praticamente todos os estudos relataram quantidade insuficiente de amostras para treinamento, visto a metodologia utilizada para esta fase requerer quantidades enormes de dados, a fim de aumentar a acurácia final do modelo. Os dados em questão, sendo imagens radiográficas, tem sua disponibilidade limitada por questões éticas e legais relativas aos pacientes e às Instituições. A falta de padronização no tratamento e processamento das imagens utilizadas nos diversos estudos provavelmente também teve influência nos resultados obtidos, dificultando a comparação e generalização destes resultados.

A existência de diversas limitações nos estudos parece mostrar que, de forma geral, a investigação da aplicabilidade clínica de modelos de IA está em seus primórdios. Apesar destas limitações, mais da metade dos trabalhos analisados mostrou acurácia dos modelos acima de 80%, e diversos estudos apresentaram

sugestões ou possíveis aperfeiçoamentos a serem utilizados em pesquisas futuras, de forma a reduzir o impacto destas limitações nos resultados.

5 CONSIDERAÇÕES FINAIS

A utilização de modelos de IA aplicados à análise de imagens radiográficas odontológicas vem sendo objeto de pesquisa em diversas áreas. Observa-se um esforço consistente no desenvolvimento e aperfeiçoamento de sistemas baseados principalmente em aprendizado profundo, que foi citado como o método de aprendizado de máquina mais adequado para análise de imagens.

O crescente número de estudos evidencia a necessidade de disponibilizar sistemas que possam auxiliar o dentista no ambiente clínico, automatizando tarefas tediosas e que requerem muito tempo para sua execução, ao mesmo tempo que fornecem dados confiáveis de diagnóstico, podendo, assim, complementar o diagnóstico feito pelo profissional.

Nesse sentido, a avaliação dos trabalhos incluídos nesta revisão permitiu inferir que a utilização de modelos de IA aplicados às imagens radiográficas odontológicas mostra, em todas as áreas pesquisadas, resultados promissores, com desempenhos muito semelhantes ou mesmo superiores, muitas vezes, ao desempenho dos profissionais humanos. A aplicação para localização automática de pontos cefalométricos e para detecção e classificação automáticas de lesões odontogênicas, em especial, parece estar em fase relativamente madura. No futuro, pode ser possível a utilização destes modelos como auxiliares no planejamento e diagnóstico quando aplicados aos exames de imagem odontológicos. Esse nível de desempenho, que deve ter precisão próxima a 100%, ainda não foi alcançado. Para a legitimização de sua efetiva utilização como parte do fluxo de trabalho na clínica, um longo caminho ainda deve ser percorrido, e diversas limitações apresentadas pelos trabalhos experimentais devem ser solucionadas como parte desse processo.

No aprendizado de máquina, a falta de grandes conjuntos de dados para treinamento ainda é uma limitação comum a todos os estudos, limitando o desenvolvimento de aplicativos auxiliares de diagnóstico. Além da quantidade, é provável que a qualidade das imagens apresente grande variação entre os diversos trabalhos, tornando muito difícil a comparação objetiva direta de resultados. Outro obstáculo comum é o pré-processamento manual das imagens para treinamento, o que despense muito tempo e esforço em cada estudo.

Ainda, em face aos crescentes avanços em reconhecimento de imagem relacionados ao aprendizado profundo, pode surgir o questionamento quanto ao papel do radiologista como especialista nas próximas décadas. Entretanto, deve-se ter em mente as limitações regulatórias impostas por órgãos competentes ao uso da IA no cenário clínico, que restringem seu papel como ferramentas auxiliares de diagnóstico e planejamento, sempre sob supervisão do dentista. Outro questionamento diz respeito à questão essencial de quem deveria assumir a responsabilidade pelos resultados diagnósticos, caso os modelos atingissem precisão compatível com uso clínico. Além disso, dúvidas envolvendo o impacto das previsões dos modelos de IA sobre a conduta clínica dos dentistas e o grau de confiabilidade e aceitação que os pacientes creditariam a esses sistemas, caso utilizados isoladamente no diagnóstico por imagens, sugerem outras linhas de investigação a serem mais bem exploradas no futuro, além das avaliações de aplicabilidade e de desempenho dos modelos já em curso.

REFERÊNCIAS

- ARIJI, Y.; YANASHITA, Y.; KUTSUNA, S. *et al.* Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. **Oral Surg Oral Med Oral Pathol Oral Radiol**, v. 128, n. 4, p. 424-430, 2019.
- BOREAK, N. Effectiveness of artificial intelligence applications designed for endodontic diagnosis, decision-making, and prediction of prognosis: a systematic review. **J Contemp Dent Pract**, v. 21, n. 8, 2020.
- CANTU, G.A.; GEHRUNG, S.; KROIS, J. *et al.* Detecting caries lesions of diferente radiographic extension on bitewings using deep learning. **J Dent**, v.100, 103425, 2020.
- CHANG, H-J.; LEE, S-J.; YONG, T-H. *et al.* Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis. **Nature, Sci Rep**, v. 10, 2020.
- CHEN, H.; ZHANG, K.; LYU, P. *et al.* A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. **Nature, Sci Rep**, v .9, n. 01, 2019.
- CHEN, S.; LI, G.; WU, T.-H. *et al.* Machine learning in orthodontics: introducing a 3D auto-segmentation and auto-landmark finder of CBCT images to assess maxillary constriction in unilateral impacted canine patients. **Angle Orthod**, v. 90, n.1, p.77-84, 2020.
- CHEN, Y-W.; STANLEY, K.; ATT, W. Artificial intelligence in dentistry: current applications and future perspectives. **Quint Int**, v. 51, n. 3, p.248-257, 2020.
- CHEN, Y-W.; STANLEY, K.; ATT, W. Corrigendum: Artificial intelligence in dentistry: current applications and future perspectives. **Quint Int**, v. 51, n. 5, p. 430, 2020.
- DOT, G.; RAFFLENBEUL, F.; ARBOTTO, M. *et al.* Accuracy and reliability of automatic three-dimensional cephalometric landmarking. **Int J Oral Maxillofac Surg**, v. 49, 2020.
- ED-DHAHRAOUY, M.; RIRI, H.; EZZAHMOULY, M. *et al.* A new methodology for automatic detection of reference points in 3D cephalometry: a pilot study. **Int Orthod**, v. 16, n. 2, p. 328-337, 2018.
- EKERT, T.; KROIS, J.; MEINHOLD, L. *et al.* Deep learning for the radiographic detection of apical lesions. **J Endod**, v. 45, n. 7, 2019.
- ENDRES, M.G.; HILLEN, F.; SALLOUMIS, M. *et al.* Development of a deep learning algorithm for periapical disease detection in dental radiographs. **Diagnostics**, v. 10, n. 430, 2020.

- FUKUDA, M.; ARIJI, Y.; KISE, Y. *et al.* Comparison of 3 deep learning neural networks for classifying the relationship between the mandibular third molar and the mandibular canal on panoramic radiographs. **Oral Surg Oral Med Oral Pathol Oral Radiol**, v. 130, n. 3, p. 336-343, 2020.
- FUKUDA, M.; INAMOTO, K.; SHIBATA, N. *et al.* Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography. **Oral Radiol**, v. 36, n. 4, p. 337-343, 2020.
- GEETHA, V.; APRAMEYA, K.S.; HINDUJA, D.M. Dental caries diagnosis in digital radiographs using back-propagation neural network. **Health Inf Sci Syst**, v. 8, n. 8, 2020.
- GRISCHKE, J.; JOHANNSMEIERS, L.; EICH, L. *et al.* Dentronics: towards robotics and artificial intelligence in dentistry. **Dent Materials**, v. 36, p. 765-778, 2020.
- HIRAIWA, T.; ARIJI, Y.; FUKUDA, M. *et al.* A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography. **Dentomaxillofac Radiol**, v. 48, n. 3, 2019.
- HUNG, K.; MONTALVAO, C.; TANAKA, R. *et al.* The use and performance of artificial intelligence applications in dental and maxillofacial radiology: a systematic review. **Dentomaxillofac Radiol**, v. 49, n. 1, 2020.
- HWANG, H.W.; PARK, J.H.; MOON, J.H. *et al.* Automated identification of cephalometric landmarks: Part 2 – Might it be better than human? **Angle Orthod**, v. 90, n. 1, 2020.
- HWANG, J.J.; JUNG, Y.H.; CHO, B.H. *et al.* An overview of deep learning in the field of dentistry. **Imaging Sci Dent**, v. 49, p. 1-7, 2019.
- JASKARI, J.; SAHLSTEN, J.; JARNSTEDT, J. *et al.* Deep learning method for mandibular canal segmentation in dental cone beam computed tomography volumes. **Sci Rep**, v. 10, n. 1, 2020.
- KATS, L.; VERED, M.; ZLOTOGORSKI-HURVITZ, A. *et al.* Atherosclerotic carotid plaque on panoramic radiographs: neural network detection. **Int J Comput Dent**, v. 22, n. 02, p. 163-169, 2019.
- KIM, J.; LEE, H-S.; SONG, I-S. *et al.* DeNTNet: deep neural transfer network for the detection of periodontal bone loss using panoramic dental radiographs. **Sci Rep**, v. 9, n. 1, p. 17615-17622, 2019.
- KIM, J.E.; NAM, N.E.; SHIM, J.S. *et al.* Transfer learning via deep neural networks for implant fixture system classification using periapical radiographs. **J Clin Med**, v. 09, 2020.
- KIM, Y.; LEE, K.; SUNWOO, L. *et al.* Deep learning in diagnosis of maxillary sinusitis using conventional radiography. **Invest Radiol**, v. 54, n. 1, p. 07-15, 2019.

KROIS, J.; EKERT, T.; MEINHOLD, L. *et al.* Deep learning for the radiographic detection of periodontal bone loss. *Nature, Sci Rep*, v. 9, 2019.

KUNZ, F.; STELLZIG-EISENHAUER, A.; ZEMAN, F. *et al.* Artificial intelligence in orthodontics. *J Orofac Orthop*, v. 81, n. 1, p. 52-68, 2020.

KUWADA, C.; ARIJI, Y.; FUKUDA, M. *et al.* Deep learning systems for detecting and classifying the presence of impacted supernumerary teeth in the maxillary incisor region on panoramic radiographs. *Oral Surg Oral Med Oral Pathol Oral Radiol*, v. 130, n. 4, p. 464-469, 2020.

KWAK, G.H.; KWAK, E-J.; SONG, J.M. *et al.* Automatic mandibular canal detection using a deep convolutional neural network. *Sci Rep*, v. 10, n. 1, p.5711-5718, 2020.

KWON, O.; YONG, T.H.; KANG, S.R. *et al.* Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolutional neural network. *Dentomaxillofac Radiol*, v. 49, n. 8, 2020.

LEE, J-H.; JEONG, S-N. Efficacy of deep convolutional neural network algorithm for the identification and classification of dental implant systems, using panoramic and periapical radiographs. A pilot study. *Medicine*, vol. 99, n.26, 2020.

LEE, J-H.; KIM, D-H.; JEONG, S-N. Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network. *Oral Diseases*, v. 26, n. 1, p. 152-158, 2020.

LEE, J-H.; KIM, D-H.; JEONG, S-N. *et al.* Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J DENT*, v. 77, p. 106-111, 2018.

LEE, J-H.; KIM, D-H.; JEONG, S-N. *et al.* Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J Period Implant Sci*, v. 48, n. 2, p. 114-123, 2018.

LEE, J-H.; KIM, Y-T.; LEE, J-B. *et al.* A performance comparison between automated deep learning and dental professionals in classification of dental implant systems from dental imaging: a multi-center study. *Diagnostics*, v. 10, n. 11, 2020.

LEE, J-S.; ADHIKARI, S.; LIU, L. *et al.* Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminar study. *Dentomaxillofac Radiol*, v. 48, n. 01, 2019.

LEE, K.S.; JUNG, S.K.; RYU, J.J. *et al.* Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs. *J Clin Med*, v. 9, n. 2, 2020.

LEE, K.S.; KWAK, H.J.; OH, J.M. *et al.* Automated detection of TMJ osteoarthritis based on artificial intelligence. *J Dent Res*, v. 99, n. 12, p. 1363-1367, 2020.

LEITE, A.F.; VASCONCELOS, K.F.; WILLEMS, H. *et al.* Radiomics and machine learning in oral healthcare. *Proteomics Clin Appl*, v. 14, 2020.

MONTÚFAR, J.; ROMERO, M.; SCOUGALL-VILCHIS, R.J. *et al.* Automatic 3-dimensional cephalometric landmarking based on active shape models in related projections. **Am J Orthod Dentofacial Orthop**, v. 153, p. 449-458, 2018.

MONTÚFAR, J.; ROMERO, M.; SCOUGALL-VILCHIS, R.J. *et al.* Hybrid approach for automatic cephalometric landmark annotation on cone-beam computed tomography volumes. **Am J Orthod Dentofacial Orthop**, v. 154, p. 140-150, 2018.

MURATA, M.; ARIJI, Y.; OHASHI, Y. *et al.* Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography. **Oral Radiol**, v. 35, n. 3, p. 301-307, 2019.

NAKAMOTO, T.; HATSUTA, S.; YAGI, S. *et al.* Computer-aided diagnosis system for osteoporosis based on quantitative evaluation of mandibular lower border porosity using panoramic radiographs. **Dentomaxillofac Radiol**, v. 49, n. 4, 2020.

NEELAPU, B.C.; KHARBANDA, O.P.; SARDANA, V. *et al.* Automatic localization of three-dimensional cephalometric landmarks on CBCT images by extracting symmetry features of the skull. **Dentomaxillofac Radiol**, v. 47, 2018.

NICHOLS, J.A.; CHAN, H.W.H.; BAKER, M.A.B. Machine learning: applications of artificial intelligence to imaging and diagnosis. **Biophysical Reviews**, v. 11, p. 111-118, 2019.

ORHAN, K.; BAYRAKDAR, I.S.; EZHOV, M. *et al.* Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. **Int Endod J**, v. 53, n. 5, p. 680-689, 2020.

PARK, J.H.; HWANG, H.W.; MOON, J.H. *et al.* Automated identification of cephalometric landmarks: Part 1 – Comparisons between the latest deep-learning methods YOLOV3 and SSD. **Angle Orthod**, v. 89, n. 6, 2019.

PARK, W.J.; PARK, J-B. History and application of artificial neural networks in dentistry. **Eur J Dent**, v. 12, n. 4, p. 594-601, 2018.

PATIL, A.; KULKARNI, V.; BHISE, A. Algorithmic analysis for dental caries detection using an adaptive neural network architecture. **Heliyon**, v. 5, n. 5, 2019.

POEDJIASTOETI, W.; SUEBNUKARN, S. Application of convolutional neural network in the diagnosis of jaw tumors. **Healthc Inform Res**, v. 24, n. 3, 2018.

PRADOS-PRIVADO, M.; VILLALÓN, J.G.; MARTINEZ-MARTINEZ, C.H. *et al.* Dental caries diagnosis and detection using neural networks: a systematic review. **J Clin Med**, v. 9, n. 11, 2020.

SAID, M.H.; LE ROUX, M.K.; CATHERINE, J.H. *et al.* Development of an artificial intelligence model to identify a dental implant from a radiograph. **Int J Oral Maxillofac Implants**, v. 35, n. 6, 2020.

SCHWENDICKE, F.; SAMEK, W.; KROIS, J. Artificial intelligence in dentistry: chances and challenges. **J Dent Res**, v. 99, n. 7, p. 769-774, 2020.

- SETZER, F.; SHI, K.; ZHANG, Z. *et al.* Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam tomographic images. **J Endod**, v. 46, n. 7, 2020.
- SILVA, V.K.S.; VIEIRA, W.A.; BERNARDINO, I.M. *et al.* Accuracy of computer-assisted image analysis in the diagnosis of maxillofacial radiolucent lesions: asystematic review and meta-analysis. **Dentomaxillofac Radiol**, v. 49, n. 3, 2020.
- SUKEGAWA, S.; YOSHII, K.; HARA, T. *et al.* Deep neural networks for dental implant system classification. **Biomolecules**, v. 10, 2020.
- TANG, A.; TAM, R.; CADRIN-CHÉNEVERT, A. *et al.* Canadian Association of Radiologists White paper on artificial intelligence in radiology. **Can Assoc Radiol J**, v. 69, p. 120-135, 2018.
- THANATHORNWONG B.; SUEBNUKARN, S. Automatic detection of periodontal compromised teeth in digital panoramic radiographs using faster regional convolutional neural networks. **Imaging Sci Dent**, v. 50, p. 169-174, 2020.
- TUZOFF, D.; TUZOVA, L.; BORNSTEIN, M. *et al.* Tooth detection and numbering in panoramic radiographs using convolutional neural networks. **Dentomaxillofac Radiol**, v. 48, n. 4, 2019.
- VINAYAHALINGAM, S.; XI, T.; BERGÉ, S. *et al.* Automated detection of third molars and mandibular nerve by deep learning. *Nature*, **Sci Rep**, v. 9, n. 1, 2019.
- YANG, H.; JO, E.; KIM, H.J. *et al.* Deep learning for automated detection of cyst and tumors of the jaw in panoramic radiographs. **J Clin Med**, v. 9, 2020.
- YUN, H.S.; JANG, T.J.; LEE, S.M. *et al.* Learning-based local-to-global landmark annotation for automatic 3D cephalometry. **Phys Med Biol**, v. 65, 2020.
- ZHANG, K.; WU, J.; CHEN, H. *et al.* An effective teeth recognition method using label tree with cascade network structure. **Comp Med Imag Graph**, v. 68, p. 61-70, 2018.

ANEXO I – TABELA I – CARACTERÍSTICAS E PERFORMANCE DOS MODELOS DE IA PROPOSTOS NOS ESTUDOS INCLUÍDOS – ORTODONTIA

AUTOR/ANO	APLICABILIDADE	MODALIDADE DE IMAGEM	TÉCNICA DE IA	Nº DE IMAGENS UTILIZADAS (TESTE)	PADRÃO DE REFERÊNCIA	RESULTADOS
Ed-Dhahraouy <i>et al.</i> (2018)	Localização automática de 21 pontos cefalométricos	TCFC	CNN específica usando linguagem C++	n=5	Localização manual em cortes multiplanares, originando imagem 3D	Erro médio global: 2,32mm; Desvio Padrão de 1,11. Maior incidência de erro ocorreu na localização dos forames infra e supraorbitais.
Montúfar <i>et al.</i> (2018)a	Localização automática de 18 pontos cefalométricos	TCFC FoV amplo	Modelo de formatação ativa (ASM)	n=24	Localização manual dos pontos por especialista	Erro médio global: 3,64mm; desvio padrão de 1,43. 12 dos 18 pontos foram reproduzidos dentro de um desvio padrão de até 1mm. Maior incidência de erro ocorreu na localização do Porio e Sela (baixa definição do volume, variação anatômica, formato irregular). Requer ajustes manuais nestes pontos.
Montúfar <i>et al.</i> (2018)b	Localização de 18 pontos cefalométricos	TCFC FoV amplo	Modelo de formatação ativa	n=24	Média dos valores da localização manual dos pontos por 2 especialistas nos cortes sagital e coronal, reconstruídos em 3D	Erro médio: 2,51mm; desvio padrão de 1,6. Tempo de processamento de imagem: 49s. Técnica permite recortar subvolumes da imagem para ajuste mais preciso do ponto.
Neelapu <i>et al.</i> (2018)	Localização automática de 20 pontos cefalométricos	TCFC FoV amplo	Algoritmo de combinação de template	n=30	Média dos valores de localização, manual, feita por 3 especialistas	Erro médio global: 1,88mm, desvio padrão de 1,1mm. Maior incidência de erro: gônio, condílio, R1 e sela.

Park <i>et al.</i> (2019)	Identificação automática de 80 pontos cefalométricos – comparação de 2 modelos	Radiografias cefalométricas 2D	CNN (YOLOv3)	n=283	Modelo YOLOv3 x modelo SSD	O modelo YOLOv3 apresentou 5% mais acurácia na detecção dos pontos cefalométricos do que o modelo SSD. Tempo para processamento de cada imagem: entre 0,05s e 2,89s. Os autores consideram o modelo com potencial uso na prática clínica.
Kunz <i>et al.</i> (2020)	Localização totalmente automatizada de 18 pontos cefalométricos	Radiografias cefalométricas digitais (2D)	CNN de código aberto (Keras & Google Tensorflow)	n=50	Valores médios de localização de cada ponto por 12 especialistas e pós-graduandos em ortodontia	Diferença de 0,37° para parâmetros angulares; <0,2mm para parâmetros lineares; <0,25% para parâmetros de proporção facial. CNN mostrou precisão comparável aos examinadores experientes. Análise automatizada efetuada em uma fração de segundo, mesmo em computador pessoal padrão.
Yun <i>et al.</i> (2020)	Localização automática de 93 pontos cefalométricos baseada em ML	TC multislice	CNN com autoencoder variacional (VAE) – estrutura de aprendizagem multi estágio	n=255	Marcações realizadas por 1 dentista com experiência de 20 anos em TC 3D	Erro médio global: 3,63 mm. O método proposto apresentou desempenho relativamente alto, mas o nível de erro não atendeu à exigência de aplicação clínica imediata (menos de 2 mm de níveis de erro). A estrutura de aprendizagem multiestágio objetiva lidar com o desafio de uma pequena quantidade de dados ao aprender recursos 3D a partir de dados de TC 3D.

Hwang <i>et al.</i> (2020)	Detecção automática de 80 pontos cefalométricos	Radiografias cefalométricas 2D	CNN (YOLOv3)	n=283	Examinadores humanos	Erro de detecção menor do que 0.9mm. Não houve diferença estatisticamente significativa entre o modelo testado e os examinadores humanos na localização dos pontos cefalométricos.
Chen <i>et al.</i> (2020)	Auto segmentação da maxila e localização automática de 3 pontos cefalométricos em casos de canino impactado unilateral	TCFC	Learning-based multi-source IntegratioN framework for Segmentatio n (LINKS) - algoritmo tipo ML	Grupo teste n=30 (com caninos impactados) Grupo controle n=30 (sem caninos impactados) Amostra de treinamento n=30	Segmentação e localização manuais	Autossegmentação por reconhecimento automático dos limites anatômicos do volume de interesse, inclusive imagens de baixa qualidade, sem necessidade de pós processamento manual. Tempo de segmentação: 15 min. Localização dos pontos cefalométricos: diferença mínima entre os grupos de 2 voxels no plano sagital mediano.

ANEXO II – TABELA II – MÉDIA DOS INDICES DE ERRO NA DETECÇÃO DE PONTOS CEFALOMÉTRICOS NOS ESTUDOS CLÍNICOS REVISADOS – ÁREA DE ORTODONTIA

AUTOR/ANO	ERRO (MÉDIA)
Ed-Dhahraouy <i>et al.</i> (2018)	2,32mm
Montúfar <i>et al.</i> (2018)	3,64mm
Montúfar <i>et al.</i> (2018)	2,51mm
Neelapu <i>et al.</i> (2018)	1,88mm
Yun <i>et al.</i> (2020)	3,63mm

**ANEXO III – TABELA III – CARACTERÍSTICAS E PERFORMANCE DOS MODELOS DE IA PROPOSTOS NOS ESTUDOS
INCLUÍDOS – CARIOLOGIA**

AUTOR/ANO	APLICABILIDADE	MODALIDADE DE IMAGEM	TÉCNICA DE IA	AMOSTRA	PADRÃO DE REFERÊNCIA	RESULTADOS		
						acurácia	sensibilidade	especificidade
Lee et al., 2018	Utilização de um algoritmo CNN DL para diagnóstico de cárie (avalia a A, S e E do modelo na detecção de cáries em imagens selecionadas e confirmadas para cáries)	Radiografias periapicais padronizadas	CNN GoogLeNet Inception v3 pré-treinada	n=600	Quatro dentistas experientes; selecionadas as imagens com 100% de acordo no diagnóstico	82%	81%	83%
						OBS: Considera o CNN com bom potencial para auxiliar o dentista no diagnóstico de cáries de esmalte e de dentina. Ressalta importância dos exames clínicos, além dos radiográficos, para o diagnóstico e do diagnóstico precoce		
Patil et al., 2019	Uso de rede neural adaptiva para detecção de cárie, comparada a classificadores convencionais	Não especificado. Não há fotografias das imagens.	Modelo MPCA-ADA	n=120	Outros três algoritmos classificadores convencionais (PCA-ADA, LDA-ADA e ICA-ADA)	Até 81% melhor	Até 90% melhor	Até 87% melhor
Cantu et al., 2020	Detecção de lesões de cárie de diferentes extensões usando modelo de aprendizado profundo	Radiografias interproximais	CNN (U-Net)	n= 141	Análise das radiografias por quatro dentistas experientes	0.80 CNN 0.71 dentistas	0.75 CNN 0.36 dentistas	0.83 CNN 0.91 dentistas
						OBS: CNN sobrepujou os dentistas na detecção de cáries em diferentes profundidades, especialmente nas iniciais. Ressalta a importância do diagnóstico precoce.		

Geetha et al., 2020	Detecção de lesões de cárie usando uma rede neural de retro propagação	Radiografias periapicais	BPNN (Back Propagation Neural Network	n=105	Outros parâmetros para técnicas de tratamento de imagem, outros classificadores para diagnóstico de cárie dental	97,1%	-	-
						Acurácia de 97% sobre outros métodos de IA, como SVM, KNN, Random forest.		

ANEXO IV – TABELA IV – CARACTERÍSTICAS E PERFORMANCE DOS MODELOS DE IA PROPOSTOS NOS ESTUDOS INCLUÍDOS – ENDODONTIA

AUTOR/ANO	APLICABILIDADE	MODALIDADE DE IMAGEM	TÉCNICA DE IA	QTDADE AMOSTRA (TESTE)	PADRÃO DE REFERÊNCIA	RESULTADOS/CONCLUSÕES
Ekert <i>et al.</i> (2019)	Detecção automática de lesões apicais	Radiografias panorâmicas (segmentadas para dentes individuais)	CNN	n=160	Detecção por 6 examinadores experientes independentes	Quando a taxa de concordância geral entre examinadores foi de 54,6%: a AUC* da CNN foi de 0.85 , especificidade de 0.87 e sensibilidade de 0.65. Sensibilidade limitada devido à baixa prevalência de lesões nas amostras. Em caso de 100% de concordância dos examinadores: AUC* de 0.95 , sensibilidade de 0.74 e especificidade de 0.94. O modelo necessita melhorar sensibilidade.
Hiraiwa <i>et al.</i> (2019)	Avaliação da morfologia radicular do 1º molar inferior (presença de 2 raízes distais) por 2 sistemas de IA	Radiografias panorâmicas (segmentação para 1ºMI)	CNN (AlexNet x GoogleNet)	n=760	Detecção da raiz distal extra por dois radiologistas com mais de 20 anos de experiência	Nas TCFC: presença de raiz D extra em 21,4% dos dentes. Acurácia do AlexNet: 87,4%. Acurácia do GoogleNet: 85,3%. (sem diferença significativa entre eles). Tempo de teste: 9s e 11s. Acurácia dos radiologistas: 81,2%. Desempenho dos modelos foi considerado igual ou superior ao dos radiologistas. Futuramente, os autores pretendem testar os modelos em radiografias periapicais.
Endres <i>et al.</i> (2020)	Detecção automática de lesões radiolucidas periapicais	Radiografias panorâmicas	CNN	n=102	Detecção das lesões por 24 cirurgiões bucomaxilofaciais (padrão ouro: análise de 197 panorâmicas por 1 CBMF + confirmação com testes clínicos)	Cirurgiões: em média, 31% falsos positivos; 49% das lesões não foram diagnosticadas. Não houve efeito significativo do tempo de experiência dos examinadores em relação aos resultados. O modelo de IA mostrou melhor desempenho diagnóstico que 58% dos dentistas, com score F1*=0.58 (Cirurgiões: PPV=0.69; TPR=0.51. CNN: PPV=0.67; TPR=0.51)

Orhan <i>et al.</i> (2020)	Detecção automática de lesões apicais (com medição de volume)	TCFC	CNN (Diagnocat)	n=153	Média dos valores de localização, manual, feita por 3 especialistas	Índice de acurácia do sistema testado: 92,8% (detectou 142 de um total de 153 lesões). TPR=0.89; PPV=0.95; F1=0.93. Medição de volume das lesões: equivalente à medição manual. Não houve diferença estatisticamente significativa entre CNN e humanos no desempenho.
Setzer <i>et al.</i> (2020)	Segmentação de imagens e detecção automática de lesões apicais	TCFC (FoV restrito)	CNN (U-Net)	n=20 imagens com 61 raízes	Segmentação e análise das imagens por 1 radiologista, 1 endodontista e 1 formando de odontologia completando programa em radiologia (índice de concordância: 93,1%).	Resultados: sensibilidade=0.93; especificidade=0.88, PPV= 0.87. Estes índices foram considerados excelentes pelos autores. Índice DICE cumulativo para todas as lesões verdadeiras: 0.67.

*AUC= Area Under the ROC Curve. A AUC varia em valor de 0 a 1. Um modelo cujas previsões estão 100% erradas tem um AUC de 0,0; aquele cujas previsões são 100% corretas tem um AUC de 1,0. *F1 Score= média harmônica da medida de precisão (valor preditivo positivo – PPV) e da sensibilidade (taxa de positivos verdadeiros – TPR). F1=1 (precisão e sensibilidade de 100%). F1=0 (precisão e sensibilidade 0).

ANEXO V – TABELA V – DESEMPENHO DOS MODELOS DE IA NA DETECÇÃO DE LESÕES RADIOLÚCIDAS PERIAPICAIS EM IMAGENS RADIOGRÁFICAS

ESTUDO	TPR (sensibilidade)	PPV (precisão)
Ekert <i>et al.</i> (2019)	0.65	0.49
Endres <i>et al.</i> (2020)	0.51	0.67
Orhan <i>et al.</i> (2020)	0.89	0.95
Setzer <i>et al.</i> (2020)	0.93	0.87

**ANEXO VI – TABELA VI – CARACTERÍSTICAS E PERFORMANCE DOS MODELOS DE IA PROPOSTOS NOS ESTUDOS
INCLUÍDOS – PERIODONTIA**

AUTOR/ ANO	APLICABILIDADE	MODALIDADE DE IMAGEM	TÉCNICA DE IA	Nº DE IMAGENS UTILIZADAS (TESTE)	PADRÃO DE REFERÊNCIA	RESULTADOS
Lee <i>et al.</i> (2018)	Diagnóstico e predição automáticos de dentes comprometidos periodontalmente	Radiografias periapicais	CNN	n=348	Diagnóstico e predição por 3 periodontistas calibrados (concordância inter- examinadores de 100%)	Acurácia do CNN no diagnóstico: para pré-molares 81% e para molares 76,7%. Acurácia de predição de exodontia em dentes clinicamente diagnosticados com doença periodontal severa: 82,8% para pré-molares e 73,4% para molares. Sem diferença estatisticamente significativa entre o modelo e os periodontistas.
Kim <i>et al.</i> (2019)	Diagnóstico automático de perda óssea periodontal	Radiografias panorâmicas	CNN	n=800	Diagnóstico por 5 dentistas clínicos experientes	Desempenho do modelo: F1 de 0.75, AUC de 0.95, sensibilidade de 0.77, especificidade de 0.95, PPV de 0.73 e NPV de 0.96. Desempenho dos especialistas: F1 de 0.69, AUC de 0.85, sensibilidade de 0.78, especificidade de 0.92, PPV de 0.62 e NPV de 0.96. Modelo com desempenho superior ao dos especialistas.
Krois <i>et al.</i> (2019)	Diagnóstico automático de perda óssea periodontal	Radiografias panorâmicas	CNN	n=1750 imagens para treinamento e validação, sem grupo de teste	Diagnóstico por 6 dentistas clínicos experientes (índice de concordância inter- examinadores de 50%)	Desempenho do modelo: acurácia de 0.81, sensibilidade de 0.81, e especificidade de 0.81. Desempenho dos especialistas: acurácia de 0.76, sensibilidade de 0.92 e especificidade de 0.63. Quanto à acurácia, não houve diferença estatisticamente significativa. Consideram que o modelo testado teve desempenho ao menos similar ao dos especialistas, mas não superior.

Chang <i>et al.</i> (2020)	Diagnóstico e estagiamento automáticos de doença periodontal	Radiografias panorâmicas	Aprendizado profundo híbrido	n=52	Diagnóstico e estagiamento de perda óssea por 3 radiologistas	Índices de acurácia diagnóstica para o modelo: 0.93 em relação à perda óssea periodontal, 0.91 em relação à junção cimento-esmalte e 0.91 para o dente analisado. PCC global de 0.73, indicando forte correlação entre os índices dos radiologistas e os do modelo, e um ICC global de 0.91, demonstrando excelente confiabilidade do desempenho diagnóstico do modelo para perda óssea periodontal.
Thanathornwong e Suebnukarn (2020)	Identificação automática de dentes comprometidos periodontalmente	Radiografias panorâmicas	Faster R-CNN	n=100 (conjunto total de imagens antes do aumento)	Diagnóstico por 3 periodontistas experientes	Índice médio de precisão de 0.81, índice de revocação médio (recall) de 0.80, sensibilidade de 0.84, especificidade de 0.88 e índice F-1 de 0.81. O modelo demonstrou habilidade de detecção satisfatória para dentes com comprometimento periodontal.

**ANEXO VII – TABELA VII – CARACTERÍSTICAS E PERFORMANCE DOS MODELOS DE IA PROPOSTOS NOS ESTUDOS
INCLUÍDOS – IMPLANTODONTIA**

AUTOR/ ANO	APLICABILIDADE	MODALIDADE DE IMAGEM	TÉCNICA DE IA	Nº DE IMAGENS UTILIZADAS	PADRÃO DE REFERÊNCIA	RESULTADOS
Kim <i>et al.</i> (2020)	Identificação automática de 4 sistemas de fixação de implante (Branemark, Dentium Implantium, Straumann Bone Level, Straumann Tissue Level)	Radiografias periapicais	CNN (teste de 5 modelos: SqueezeNet, GoogLeNet, ResNet-18, MobileNet-v2 e ResNet-50)	n=801	Não há	Todos os 5 modelos mostraram acurácia acima de 90%. Os resultados deste estudo confirmaram que redes neurais convolucionais podem classificar os quatro sistemas de fixação de implante com alta precisão, mesmo com uma rede relativamente pequena, um pequeno número de imagens e implantes de configuração semelhantes.
Lee, J.H e Jeong, S.N. (2020)	Identificação automática de 3 sistemas de implante semelhantes (TSIII AS, Superline e Straumann BLT)	Radiografias panorâmicas e periapicais	CNN (GoogLeNet Inception-v3)	n=10.770 (5.390 panorâmicas e 5.380 periapicais)	Identificação dos 3 sistemas de implante por 1 periodontista	Desempenho do modelo: AUC=0,971; sensibilidade de 95.3 e especificidade de 97.6. Desempenho do especialista: AUC=0,925; sensibilidade de 88.7 e especificidade de 87.1. O modelo mostrou resultados confiáveis e acurácia maior do que o periodontista, embora não haja diferença estatisticamente significativa quanto ao desempenho geral entre eles.
Lee <i>et al.</i> (2020)	Identificação automática de 6 sistemas de implante	Radiografias panorâmicas e periapicais	DCNN automatizado	n=11.980	Diagnóstico por 25 dentistas clínicos	Desempenho do modelo: acurácia de 0.954, sensibilidade de 0.955, e especificidade de 0.853. O tipo de radiografia não foi significativo. Consideraram que o modelo testado superou o desempenho dos especialistas.

Said <i>et al.</i> (2020)	Identificação automática de 6 modelos de implante	Radiografias panorâmicas e periapicais	CNN	n=1.206	Não há	Desempenho do modelo: acurácia de 93,8%, sensibilidade de 93,5%, especificidade de 94,2%, PPV de 92%, NPV de 91,5%.
Sukegawa <i>et al.</i> (2020)	Classificação automática de 11 sistemas de implantes	Radiografias panorâmicas	5 modelos DCNN	n=8.859	Não há	Desempenho dos modelos: todos apresentaram acurácia diagnóstica entre 0.860 e 0.935, mostrando-se altamente eficazes na identificação de 11 sistemas de implantes em panorâmicas.

**ANEXO VIII – TABELA VIII – CARACTERÍSTICAS E PERFORMANCE DOS MODELOS DE IA PROPOSTOS NOS ESTUDOS
INCLUÍDOS – PATOLOGIA**

AUTOR/ ANO	APLICABILIDADE	MODALIDADE DE IMAGEM	TÉCNICA DE IA	Nº DE IMAGENS UTILIZADAS	PADRÃO DE REFERÊNCIA	RESULTADOS
Poedjiastoeti e Suebnukarn (2018)	Identificação automática de ameloblastomas e ceratocistos odontogênicos em mandíbula	Radiografias panorâmicas	CNN	n=500	Diagnóstico das lesões por 5 especialistas bucomaxilofaciais	Desempenho do CNN e dos especialistas foi similar: CNN - sensibilidade de 81,8%, especificidade de 83,3%, acurácia de 83% . Especialistas - sensibilidade de 81,1%, especificidade de 83,2%, acurácia de 82,9%. O diferencial foi o tempo para o diagnóstico: CNN levou 38s , os especialistas precisaram de 23,1 minutos.
Ariji <i>et al.</i> (2019)	Identificação automática de 5 tipos de lesões radiolúcidas em mandíbula (ameloblastoma, ceratocisto odontogênico, cisto dentífero, cisto radicular e cisto ósseo simples)	Radiografias panorâmicas	CNN	n=285	Diagnóstico das lesões por exame histopatológico	O modelo foi capaz de detectar corretamente a presença de lesão em 88% das imagens; detectou falso positivo em 4% das imagens. Quanto à classificação, o índice variou de 13%, para ceratocisto odontogênico (que foram corretamente detectados em 100%), e 82%, para cisto dentífero (detectados em 88%). O melhor desempenho para detecção e classificação foi para cistos dentíferos. O tempo de diagnóstico foi de 13s .
Lee <i>et al.</i> (2020)	Diagnóstico de 3 tipos de lesões císticas (ceratocisto odontogênico, cisto dentífero e cisto periapical)	Radiografias panorâmicas e TCFC	CNN	n=2.126 (1.140 panorâmicas e 986 TCFC)	Diagnóstico das lesões por exame histopatológico	Ao utilizar imagens panorâmicas , a precisão diagnóstica total foi de 84,6% ; a precisão diagnóstica foi maior para cistos periapicais (87,0%) e menor para ceratocistos (81,8%). Para as TCFC , a acurácia total de diagnósticos foi de 91,4% ; a precisão diagnóstica foi maior para cistos periapicais (93,7%) e menor para ceratocistos (87,2%).

Kwon <i>et al.</i> (2020)	Diagnóstico automático de cistos e tumores odontogênicos de mandíbula e maxila	Radiografias panorâmicas	CNN (modificação do YOLOv3)	n=1.282	Diagnóstico das lesões por exame histopatológico	O desempenho global do modelo revelou acurácia de 95,6% para o diagnóstico de doença. O melhor desempenho diagnóstico foi para cistos dentígeros, com acurácia de 97,8%. A detecção de normalidade teve acurácia de 96%.
Yang <i>et al.</i> (2020)	Diagnóstico automático de cistos e tumores odontogênicos (cisto dentígero, ceratocisto odontogênico e ameloblastoma) em mandíbula e maxila	Radiografias panorâmicas	CNN (YOLOv2)	n=1.603	Diagnóstico das lesões por exame histopatológico Diagnóstico por cirurgiões bucomaxilofaciais e clínicos gerais	A acurácia diagnóstica do modelo foi de 66% , enquanto a dos especialistas variou de 64% a 66% e a dos clínicos gerais ficou entre 58% e 60%. Não houve diferença estatisticamente significativa quanto ao desempenho diagnóstico entre o modelo e os profissionais. Tempo médio para avaliação das imagens: 33,8 min para os humanos; modelo: detecção instantânea.

ANEXO IX – TABELA IX – ACURÁCIA DIAGNÓSTICA DAS CNN APLICADAS AO DIAGNÓSTICO DE LESÕES ODONTOGÊNICAS EM MANDÍBULA E MAXILA – RESUMO DOS CINCO ESTUDOS ANALISADOS

ESTUDO	EXAME DE IMAGEM	ACURÁCIA DIAGNÓSTICA	CNN X PROFISSIONAIS
Poedjiastoeti e Suebnukarn (2018)	Panorâmicas	83%	Similar
Ariji <i>et al.</i> (2019)	Panorâmicas	88%	-
Lee <i>et al.</i> (2020)	Panorâmicas TCFC	84,6% 91,4%	- -
Kwon <i>et al.</i> (2020)	Panorâmicas	95,6%	-
Yang <i>et al.</i> (2020)	Panorâmicas	66%	Similar

ANEXO X – TABELA X – CARACTERÍSTICAS E PERFORMANCE DOS MODELOS DE IA PROPOSTOS NOS ESTUDOS INCLUÍDOS – DIAGNÓSTICO RADIOLÓGICO

APLICABILIDADE	AUTOR/ANO	MODALIDADE DE IMAGEM	TÉCNICA DE IA	REFERENCIAL COMPARATIVO	RESULTADOS/CONCLUSÃO
Detecção de osteoporose	Lee <i>et al.</i> (2019)	Radiografias panorâmicas	CNN	Diagnóstico feito por radiologistas experientes	Diagnóstico de osteoporose: erosão cortical na região central da base mandibular dos lados esquerdo e direito. Desempenho do modelo: valores acurácia variaram de 92,5% a 98,5%; alta concordância entre o modelo e os especialistas. Eficaz mesmo para osteoporose inicial. Confirmou região central da base mandibular dos lados esquerdo e direito como preferenciais para análise do tecido ósseo para osteoporose. Acurácia diagnóstica entre 66% e 84%. Desempenho diagnóstico: sensibilidade 90,9%, especificidade 64,7%, acurácia 75%.
	Lee <i>et al.</i> (2020)	Radiografias panorâmicas	CNN3 VGG-16 VGG-16 TF VGG-16 TF FT	Comparação entre os 4 modelos entre si	
	Nakamoto <i>et al.</i> (2020)	Radiografias panorâmicas	Computer-aided diagnosis (CAD) system	Comparação com exame de densidade óssea lombar	
Diagnóstico de sinusite maxilar	Kim <i>et al.</i> (2019)	Radiografia Waters	Algoritmo de aprendizado profundo	Diagnóstico feito por 5 radiologistas experientes; tomografias de feixe cônico dos seios paranasais	O algoritmo de aprendizagem profunda mostrou AUC significativamente maior estatisticamente do que o radiologista. Em termos de sensibilidade e especificidade, o algoritmo foi comparável aos radiologistas. Sem diferença estatisticamente significativa entre a CNN e os radiologistas na acurácia diagnóstica.
	Murata <i>et al.</i> (2019)	Radiografias panorâmicas	AlexNet CNN	Diagnóstico por 4 profissionais; tomografias de feixe cônico dos seios paranasais	

Detecção e numeração de elementos dentários	Zhang <i>et al.</i> (2018)	Radiografias periapicais	CNN em cascata	-	Precisão de 95,8% e recall de 96,1%
	Chen <i>et al.</i> (2019)	Radiografias periapicais	Faster R-CNN	Detecção e numeração por 3 dentistas	Precisão e recall de 98% para detecção e de 91% para numeração. Desempenho diagnóstico do modelo foi similar ao de um dentista recém-formado.
	Tuzoff <i>et al.</i> (2019)	Radiografias panorâmicas	CNN		Precisão e sensibilidade para detecção de 99%, e de 99% e 98% para numeração, respectivamente. Desempenho equivalente ao dos especialistas.
	Kuwada <i>et al.</i> (2020)	Radiografias panorâmicas	AlexNet VGG-16 Detect-Net	Comparação dos 3 sistemas entre si	Objetivo: classificar dentes supranumerários impactados na região anterior da maxila em pacientes com incisivos totalmente erupcionados. O desempenho de detecção do DetectNet mostrou que o <i>recall</i> , a precisão e o índice F1 foram todos 1.0, indicando detecção perfeita.
Detecção automática de terceiros molares e canal mandibular	Vinayahalingam <i>et al.</i> (2019)	Radiografias panorâmicas	CNN (U-Net)	Segmentação manual em panorâmicas	Média do Coeficiente DICE para detecção de terceiros molares foi de 93,6% e do canal mandibular, de 80,5%. Média dos índices de sensibilidade para terceiros molares foi de 94,7% e para canal mandibular, de 84,7%. Média dos índices de especificidade para terceiros molares foi de 99,9% e para canal mandibular foi de 96,7%. Autores sugerem melhorias do algoritmo para melhorar a acurácia diagnóstica.
	Fukuda <i>et al.</i> (2020)	Radiografias panorâmicas	AlexNet, GoogLeNet e VGG-16	Comparação dos 3 sistemas entre si	
	Jaskari <i>et al.</i> (2020)	TCFC	CNN	Segmentação manual por 2 radiologistas	Os modelos tiveram melhor desempenho nos recortes de imagem menores, com valores de AUC variando de 0.88 a 0.93, enquanto os valores para os recortes maiores variaram de 0.75 a 0.93. O GoogLeNet aplicado aos

	Kwak <i>et al.</i> (2020)	TCFC (imagens 2D e 3D)	2D SegNet 2D U-Net 3D U-Net	Segmentação manual por radiologistas	recortes menores teve o melhor valor da AUC, de 0.93. Localização automática do canal mandibular em reconstruções 3D visando inserção de implantes na mandíbula. A precisão da previsão do modelo foi da ordem de 0,5 mm para cerca de 90% do comprimento do canal mandibular. Índices de acurácia entre 0.76 e 0.99, com os maiores índices para imagens 3D.
Detecção de placas ateroscleróticas carotídeas	Kats <i>et al.</i> (2019)	Radiografias panorâmicas	Faster R-CNN	Detecção por 2 radiologistas	Sensibilidade de 75%, especificidade de 80% e acurácia de 83%. Os autores consideram os resultados promissores e sugestivos de potencial para futura utilização na prática clínica, necessitando estudos adicionais para aperfeiçoamentos do modelo.
Detecção de fratura radicular vertical	Fukuda <i>et al.</i> (2020)	Radiografias panorâmicas	CNN	-	Recall de 0.75, precisão de 0.93 e escore F1 (média harmônica entre recall e precisão) de 0.83. Os autores consideram que o desempenho do modelo pareceu insuficiente para aplicação na prática clínica.
Detecção de osteoartrite em ATM	Lee <i>et al.</i> (2020)	TCFC (cortes sagitais)	SSD (deep-learning)	-	Acurácia de 0.86, precisão de 0.85, recall de 0.84 e escore F1 de 0.84. Potencial para auxiliar o radiologista no diagnóstico.