

Estimativa de Idade por meio Redes Neurais Profundas: Uma Revisão Sistemática da Literatura

Michele Faria de Oliveira

Fischer Jônatas Ferreira

Iális Cavalcante de Paula Júnior

Universidade Federal do Ceará (UFC) Universidade Federal de Itajubá (UNIFEI) Universidade Federal do Ceará (UFC)

Resumo—A estimativa de idade cronológica é o processo de determinar a faixa etária de uma pessoa, com base em características biométricas. Os métodos tradicionais de identificação da idade por meio da arcada dentária são amplamente utilizados na odontologia forense. Porém, a necessidade de análises menos subjetivas ao viés do olhar humano motivou o uso de redes neurais profundas para atender este propósito. Mesmo com diversos trabalhos nesta direção, ainda falta uma sumarização e identificação das principais abordagens descritas na literatura. Desta forma, com o objetivo de levantar o estado da arte sobre o tema e apoiar futuras pesquisas, este trabalho busca selecionar os estudos primários relatados na literatura por meio de uma revisão sistemática. Foram encontrados 85 trabalhos, dos quais 17 foram selecionados para responder as questões de pesquisa formuladas. Os principais resultados alcançados foram: (i) *ResNet* obteve o melhor ganho para classificação, e *SOS-Net* para regressão entre as demais redes neurais analisadas. (ii) Para as crianças (de 3 a 14 anos) as redes apresentaram melhores resultados em comparação às demais faixas etárias. (iii) As métricas acurácia e MAE foram as mais utilizadas para avaliar as redes neurais. Essas contribuições podem apoiar pesquisadores, desenvolvedores de ferramentas, bem como cientistas forenses e profissionais da odontologia.

Palavras-chave—Redes Neurais Profundas, Estimativa de Idade, Odontologia Forense e Revisão sistemática da Literatura

I. INTRODUÇÃO

A estimativa de idade é o processo de determinar a faixa etária de uma pessoa com base em características biométricas [1]. Hoje em dia, a pesquisa na área tem ganhado muito destaque, por conta das necessidades impostas à ciência forense [1]–[4]. Devido a questões adversas tais como acidentes aéreos, explosões e catástrofes onde se faz necessária a identificação de pessoas, a estimativa de idade pode ser uma aliada importante para ajudar na identificação das pessoas falecidas [1], [5]–[7]. Além das demandas de atividades forenses, a classificação da faixa etária por meio da arcada dentária pode ajudar em outras aplicações, tais como serviços de passaporte [4], identificação de maioridade [8], planejamento ortodôntico, arqueologia [9] e odontologia pediátrica [10].

Os métodos tradicionais de identificação da idade por meio da arcada dentária são amplamente utilizados na odontologia forense e incluem diversas técnicas baseadas na observação e análise do desenvolvimento dentário [8]. Essas abordagens geralmente envolvem a avaliação de características morfológicas, como a erupção dentária, a formação das raízes e o grau de mineralização dos dentes, sendo aplicadas tanto em indivíduos vivos quanto em restos mortais [9], [11]. Um dos

métodos mais populares e consolidados na literatura é o de Demirjian [12], que classifica o estágio de desenvolvimento de sete dentes do lado esquerdo da mandíbula, oferecendo uma pontuação que pode ser convertida em idade estimada [6]. Este método é amplamente reconhecido e aplicado em várias populações ao redor do mundo devido à sua simplicidade, reprodutibilidade e relativa precisão. No entanto, estudos recentes têm apontado limitações na sua generalização, especialmente quando aplicado a grupos populacionais distintos daqueles para os quais foi originalmente desenvolvido, o que pode comprometer sua acurácia [4]–[7], [12].

Ainda, vale destacar que outros métodos tradicionais também têm grande relevância na estimativa de idade dental, tais como os métodos de Nolla [13], Nicodemo et al. [14] e Kvaal et al. [15]. Cada uma dessas técnicas possui seus pontos fortes, como a objetividade na mensuração de estágios (Nolla), a integração de múltiplos parâmetros de crescimento (Nicodemo) e o uso de relações proporcionais entre estruturas dentárias (Kvaal), mas também compartilham limitações semelhantes às do método de Demirjian, sobretudo no que diz respeito à variabilidade populacional e à necessidade de calibração para diferentes grupos étnicos.

Diante disso, torna-se cada vez mais relevante a busca por formas menos custosas, mais objetivas e automatizadas para apoiar a estimativa de idade por meio de radiografias dentárias, como aquelas baseadas em inteligência artificial e redes neurais profundas [6], [16]. Além disso, as RNP estão sendo aplicadas em outros contextos relacionados à saúde [17]–[19]. Tais abordagens oferecem potencial para reduzir o viés humano, melhorar a precisão dos resultados e ampliar a aplicabilidade dos métodos em diferentes contextos forenses e populacionais. Oportunamente, na literatura relacionada existem relatos do uso de *Redes Neurais Profundas* (RNP) para identificação de idade por meio da arcada dentária [1], [12], [20], e ainda os trabalhos recentes têm mostrado resultados promissores do uso de redes neurais com essa finalidade específica [3], [21], [22].

No entanto, observa-se uma carência de estudos sistemáticos e replicáveis que consolidem as evidências disponíveis sobre o uso de RNP na estimativa de idade por meio de radiografias panorâmicas odontológicas [2], [10], [12], [16], [22]. Embora existam diversos estudos com diferentes arquiteturas de redes e propostas metodológicas, ainda é difícil identificar, de forma clara, quais abordagens são mais eficazes, quais faixas etárias são mais bem classificadas, quais métricas são

utilizadas e quais fatores interferem nos resultados, pelo fato da baixa replicabilidade desses estudos. Essa fragmentação do conhecimento e baixa replicabilidade dos estudos dificulta a consolidação dos resultados e a real eficácia do uso das redes neurais profundas para estimativa de idades [12].

Diante do exposto, este trabalho tem como objetivo identificar e sumarizar os estudos primários relatados na literatura sobre o uso de *Redes Neurais Profundas para Estimativa de Idade por meio de Radiografias Odontológicas Panorâmicas* (REIP). Para esse propósito, foi conduzida uma *Revisão Sistemática da Literatura* (RSL) em seis bases de dados e foram encontrados 85 estudos, destes, foram selecionados 17 trabalhos para leitura completa. Como principais resultados desta RSL foram encontradas 25 redes utilizadas para classificação (três ocorrências), regressão (14 ocorrências) e para ambos os propósitos, foram encontradas oito redes. Ainda, foram apontadas as principais características dessas redes bem como indicações da eficácia de seu uso considerando diferentes faixas etárias. Estes resultados podem apoiar pesquisadores, construtores de ferramentas baseadas em redes neurais profundas, além de profissionais da área forense.

As principais contribuições deste trabalho são:

- Uma breve visão geral sobre as características das redes neurais profundas encontradas;
- Repositório público com as redes neurais profundas encontradas neste trabalho¹.

O restante deste artigo relata o planejamento, execução e os resultados da RSL conduzida e está organizado da seguinte forma: a Seção II apresenta o planejamento da RSL demonstrando o protocolo adotado para o estudo. A Seção III descreve a execução da revisão de acordo com o protocolo e mostra o processo de seleção dos artigos relevantes. A Seção IV apresenta e discute os resultados da RSL. Ainda, a Seção V discute as principais ameaças à validade do estudo e a Seção VI discute as conclusões e trabalhos futuros.

II. PLANEJAMENTO

Na fase de planejamento, foi definido o protocolo para conduzir a RSL. As atividades realizadas nesta fase foram: (i) definir o objetivo da RSL, (ii) especificar as questões de pesquisa; (iii) selecionar as bases de dados eletrônicas para pesquisar os trabalhos; (iv) construir a *string* de busca e (v) aplicar os critérios de inclusão e exclusão.

Objetivo: Identificar e analisar REIP. Foi definido este objetivo, devido ao constante relato do uso de RNP na prática forense para estimativa de idade a partir de radiografias. É esperado que pesquisadores e usuários de RNP possam se beneficiar de tal sumarização dos trabalhos relatados neste estudo.

Questões de pesquisa: As questões de pesquisa (QP) visam investigar e sumarizar o estado da arte do uso de REIP. Para atingir esse objetivo, foram estabelecidas três questões de pesquisa de propósito geral:

QP 1 - Quais são as características dos trabalhos feitos para avaliar as RNP? Por meio dessa questão de pesquisa serão

sumarizadas as principais características dos estudos escolhidos nessa RSL. Foram buscadas informações, tais como: a disponibilidade das bases de dados utilizadas, técnicas de processamento de imagem entre outras características.

QP 2 - Quais são as principais RNP utilizadas? Essa QP foi formulada com o intuito de identificar quais as principais redes neurais profundas mencionadas na literatura sobre REIP. Por meio dessa questão de pesquisa, é esperada a sumarização das principais redes neurais profundas da literatura.

QP 3 - Quais são as principais características das RNP? Essa QP tem como intuito fazer uma sumarização das principais características das RNP. É esperado que seja apresentada uma sumarização dessas características para auxiliar os usuários dessas redes na escolha daquelas que melhor se adaptem às suas necessidades.

Bases de dados eletrônicas: As bases de dados usadas para a busca dos estudos primários foram: ACM Digital Library², PubMed³, Science Direct⁴, Scopus⁵, Springer Link⁶ e Web of Science⁷. Elas foram escolhidas porque apresentam uma grande coleção de trabalhos. Além disso, abrigam trabalhos de conferências e periódicos de grande importância para a comunidade acadêmica. Por meio de uma revisão *ad-hoc* foram selecionados alguns trabalhos para compor o estudo. Assim, foi observado quais bases de dados indexavam os trabalhos encontrados, o que também motivou a escolha delas.

String de busca: Para identificar os trabalhos relevantes sobre REIP, foi formulada uma *string* de busca para encontrar estudos primários. Com base em trabalhos relacionados na revisão *ad hoc*, foram selecionadas algumas palavras-chave para compor a *string* de busca. A Tabela I apresenta as palavras-chave escolhidas para compor a *string* de busca. Foram selecionados três tópicos de interesse (primeira coluna) e, em seguida, foram selecionadas as palavras-chave (segunda coluna). Por meio dessas palavras-chave selecionadas foi elaborada a seguinte *string* de busca.

TABLE I
PALAVRAS-CHAVE SELECIONADAS

Tópicos	Palavras-chave selecionadas
Redes neurais	Artificial Neural Networks (ANN), Deep Learning, Machine Learning, Neural Network Architectures, Supervised Learning, Unsupervised Learning, Recurrent Neural Networks (RNN) Convolutional Neural Networks (CNN), Backpropagation, Neural Network Training, Neural Network Applications, Neural Network Optimization, Feedforward Neural Networks, Neural Network Modeling, Neural Network Algorithms, Neural Network Performance
Imagens radiográficas	Panoramic Dental Radiographs, Panoramic Radiography Dental Panoramic X-rays Orthopantomogram, Panoramic Oral Radiography
Faixa etária	Age Groups, Age Classification, Age-Specific Analysis, Age Categories

Para avaliar se a *string* de busca estava recuperando trabalhos relevantes foi verificado se os trabalhos inicialmente selecionados por meio da revisão *ad hoc* poderiam ser encontrados. Foram identificados na revisão *ad hoc* seis trabalhos de interesse, sendo que quatro trabalhos também puderam ser retornados pela *string* de busca.

Crítérios de Inclusão e Exclusão: Os critérios de inclusão e exclusão permitem classificar cada estudo na revisão sis-

¹ <https://rsl-redes-neurais-profundas.github.io/rsl/>

² <http://dl.acm.org/>

³ <https://pubmed.ncbi.nlm.nih.gov/>

⁴ <https://www.sciencedirect.com/>

⁵ <https://www.scopus.com/>

⁶ <http://link.springer.com/>

⁷ <https://shre.ink/bn01>

(Artificial Neural Networks OR ANN OR Deep Learning OR Neural Networks OR Machine Learning OR Convolutional Neural Networks OR CNN) **AND** (Panoramic Dental Radiographs OR Panoramic Radiography OR Dental Panoramic X-rays OR Orthopantomogram OR Panoramic Oral Radiography) **AND** (Age Groups OR Age Classification OR Age-Specific Analysis OR Age Categories)

Fig. 1. Expressão booleana utilizada para a busca sistemática.

temática da literatura como um candidato a ser incluído ou excluído [23]. Como uma RSL pode envolver um grande número de estudos, foi limitado o escopo deste trabalho para selecionar apenas artigos em inglês publicados e revisados por pares, disponíveis em formato eletrônico, publicados em conferências e periódicos e relacionados aos termos da *string* de busca. Aqueles trabalhos considerados como literatura cinza foram removidos.

Foram incluídos estudos que: **1 - relacionados a REIP** (que usa arquiteturas de RNP para estimar idade); **2 - Faixa etária** (consideram a perspectiva de grupos etários com uso de REIP); **3 - Uso de redes neurais** (relatam o uso de RNPs, redes convolucionais, redes recorrentes, e outras arquiteturas relevantes para REIP); **4 - Estudos empíricos** (realizam análise comparativa entre diferentes modelos de REIP); **5 - Métricas** (reportam métricas de desempenho (ex: acurácia, precisão, recall, F1-score) de REIP); **6 - Publicação** (foram publicados nos últimos 5 anos).

Foram excluídos estudos que: **1 - Métodos que não usam RNP** (não focam explicitamente em métodos de REIP, mas apenas se referem à RNP como um trabalho paralelo); **2 - Estudos que não apresentam uma descrição do comportamento das RNP utilizadas** (não apresentam uma validação e caracterização das RNPs utilizadas no estudo e que não apresentam resultados empíricos sobre o uso de REIP); **3 - Uso de RNP que não são voltados para estimativa de idades** (que utilizam RNP para outras aplicações odontológicas, como segmentação ou detecção de cáries, e que não estejam focados em REIP); **4 - Estudos que não identificam idade dos pacientes** (que não consideram a perspectiva de grupos etários); **5 - Prática odontológica** (que não aplicam os modelos de RNP no contexto clínico ou que falham em relacionar os resultados à prática odontológica); **6 - Artigos curtos** (em formato de resumo sem acesso ao texto completo, ou trabalhos curtos de uma ou duas páginas); **7 - Literatura cinza** (teses, dissertações, e relatórios técnicos que não foram publicados em revistas revisadas por pares ou conferências relevantes).

Seleção inicial do estudo: Os pesquisadores autores desse trabalho leram o título, resumo e palavras-chave dos estudos recuperados e os classificaram como incluídos ou excluídos, conforme os critérios de inclusão e exclusão estabelecidos. Para este trabalho foi adotada a seguinte avaliação: um estudo foi incluído se ele satisfaz todos os critérios de inclusão e foi excluído se satisfaz ao menos um critério de exclusão.

Seleção final: Os autores desse trabalho fizeram a leitura completa dos artigos selecionados para identificar os tópicos de interesse a serem sumarizados para compor o corpo de conhecimento deste trabalho. A Tabela II apresenta quais

foram os itens analisados para sumarizar os artigos selecionados. Estes itens foram separados em três grupos: classificação do trabalho, qualificação do estudo e métricas utilizadas. Para auxiliar todo o processo de planejamento, execução, e análise dos dados reportados na RSL foi utilizada a ferramenta START [24]. Essa ferramenta auxilia todas as etapas de criação de uma RSL.

TABLE II
ITENS ANALISADOS NOS TRABALHOS SELECIONADOS

Grupos	Item analisado
Classificação do trabalho	Ano de publicação, Tipo de publicação, Evento publicado, Classificação CAPES e Número de Citações
Qualificação do estudo	1) De qual população (país) foram levantados os dados? 2) O estudo apresenta uma rede neural? 3) O estudo apresenta o número de imagens analisadas? Qual foi? 4) Qual é o grupo etário que o estudo trabalha? 5) A base de dados utilizada no artigo é pública? 6) Foi feita uma técnica de processamento nas imagens? 7) Foi utilizado um classificador? 8) O estudo apresenta uma classificação de imagens de radiografias odontológica panorâmica?
Métricas utilizadas	9) Quais foram as métricas utilizadas? 10) Caso tenha, qual foi a acurácia relatada? 11) O estudo reporta métricas de desempenho?

III. EXECUÇÃO

Esta seção apresenta a execução dos protocolos descritos na Seção II. Assim, foi aplicada a *string* de busca nas seis bases de dados para recuperar estudos primários. Com a finalidade de encontrar abordagens mais atuais e consolidadas na literatura foram incluídos apenas trabalhos publicados entre os anos de 2020 e 2024. O processo de busca foi realizado de 11 de agosto a 1º de setembro de 2024.

Com as informações dos trabalhos encontrados, extraídas no formato *BibTeX* para a ferramenta START, os critérios de inclusão e exclusão foram aplicados à filtragem dos estudos. As informações dos trabalhos tais como título, resumo e palavras-chave dos trabalhos foram analisadas diretamente na ferramenta START. Foi dada uma classificação de prioridade de leitura para cada trabalho como: muito alta, alta, baixa e muito baixa. Além da classificação: aceito ou rejeitado.

A Figura 2 apresenta as etapas e buscas dos estudos selecionados nesta RSL. Inicialmente, foi feita a busca dos trabalhos nas bases de dados, sendo que nessa etapa foram encontradas 85 publicações. A *Science Direct* foi a base que mais retornou trabalhos, com um total de 25 publicações e a ACM foi a que menos retornou, com apenas quatro trabalhos. Além disso, dois estudos identificados na Revisão *ad hoc* não estavam presentes em nenhuma das bases consultadas, mas foram incluídos manualmente para compor o conjunto de estudos analisados. Outros quatro estudos provenientes da Revisão *ad hoc* foram encontrados nas bases de dados, sendo dois na *Springer Link*, um na *Web of Science* e um na *Scopus*.

Posteriormente, o processo de seleção dos trabalhos, conforme indicado na Figura 2 foi feito em três etapas: remoção de duplicados (A), sendo que 20 artigos foram classificados como duplicados, esta etapa foi feita por meio da ferramenta START, assim, 65 trabalhos passaram para a próxima etapa (B). Após a leitura do título, resumo e palavras-chave dos trabalhos 22 artigos foram selecionados e 42 artigos foram rejeitados. Finalmente, foram lidos os textos completos dos 22

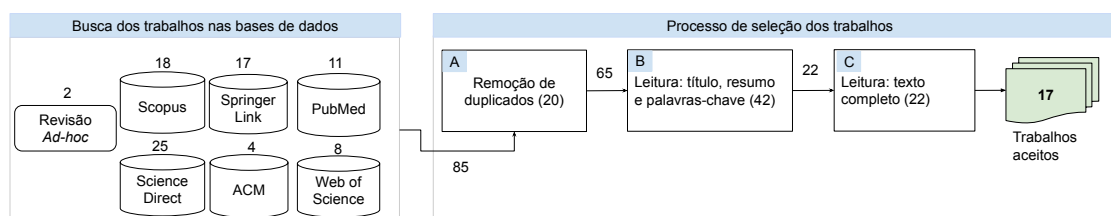


Fig. 2. Processo de busca e seleção dos trabalhos

artigos (C), dos quais 5 foram excluídos por não atenderem aos critérios de inclusão e exclusão. Dessa forma, esta RSL apresenta a sumarização de 17 trabalhos considerados relevantes para responder às questões de pesquisa levantadas na Seção II. A lista dos trabalhos selecionados nesta revisão sistemática da literatura está disponibilizada no repositório público que acompanha o pacote de replicação do estudo.

IV. ANÁLISE

Esta seção apresenta as respostas às questões de pesquisa (QP) realizadas na Seção II. Dessa forma, a Seção IV-A responde a QP sobre as principais características dos estudos selecionados, apresentando uma visão geral desses estudos. A Seção IV-B apresenta as redes neurais que os estudos selecionados utilizam. Finalmente, a Seção IV-C sumariza as principais características destas redes neurais identificadas.

A. Quais são as características dos trabalhos feitos para avaliar as RNP?

Em resposta à RQ1 formulada na Seção II, esta seção apresenta uma visão geral dos estudos primários selecionados na RSL. Com a finalidade de qualificar os trabalhos selecionados no estudo, a Figura 3 apresenta o quantitativo de trabalhos em relação ao Qualis (Figura 3(a)), população do estudo (Figura 3(b)) ano de publicação (Figura 3(c)) e do evento de publicação (Figura 3(d)). Como pode ser observado em relação à relevância dos trabalhos, a maioria dos estudos (cerca de 76%) apresenta o Qualis superior (A1, A2 e A3). Porém, para três estudos não foi possível identificar ou eles não possuem Qualis. Além disso, foi identificado para a população de quais países os estudos pertencem. A China e Coreia do Sul apresentam seis e quatro trabalhos, respectivamente. Foi selecionado apenas um estudo referente à população brasileira. Ainda, no ano de 2024 foram publicados seis trabalhos, seguido do ano de 2023 com cinco trabalhos. Como pode ser observado na Figura 3(c) a pesquisa relacionada a REIP está em ascensão. Finalmente, dos 17 trabalhos selecionados 16 (cerca de 94%) trabalhos são estudos de periódicos e um trabalho foi publicado em *workshop*. Os eventos onde os trabalhos selecionados foram publicados tem como principais destaques, os periódicos *Scientific Reports* e *BMC Oral Health* foi com três publicações cada.

Foi analisado se nos estudos selecionados utilizaram-se técnicas de pré-processamento nas imagens. A Figura 4(a) apresenta uma sumarização das técnicas de processamento das imagens como preparação para redes neurais. Como pode ser

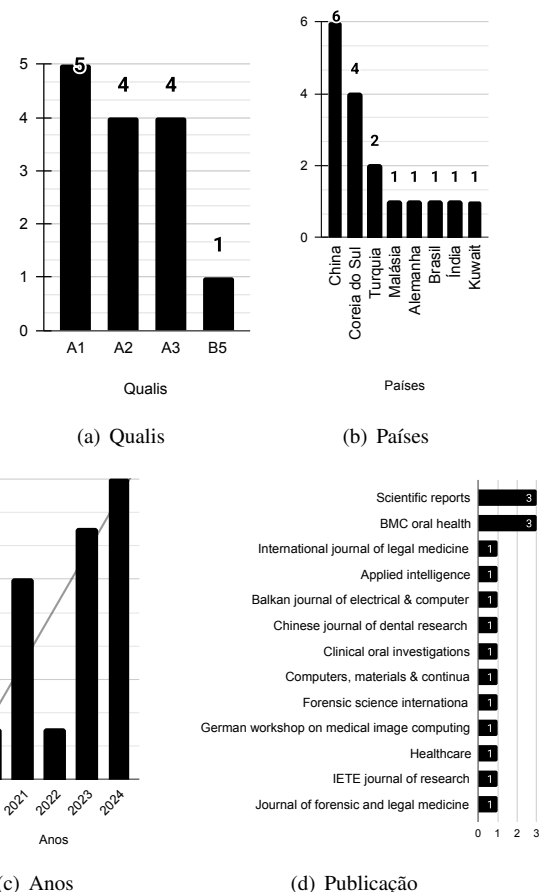
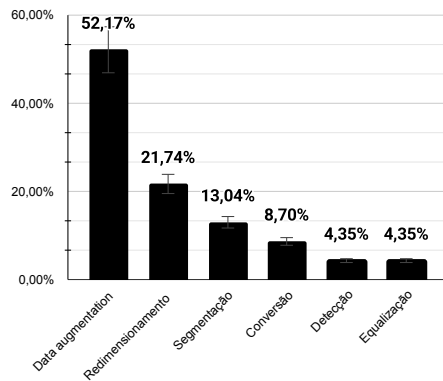


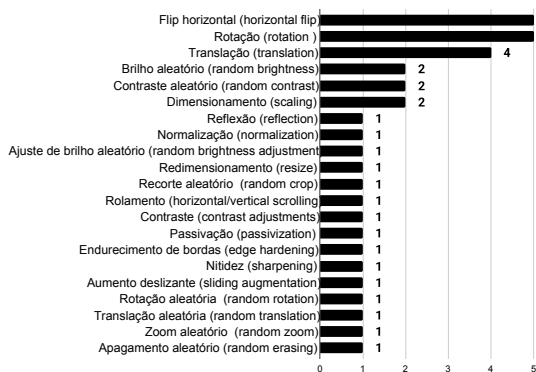
Fig. 3. Distribuição de trabalhos por Qualis e Países

observado, o processamento mais utilizado foi o *data augmentation*, cerca de 52% dos trabalhos selecionados utilizam essa técnica para preparar suas imagens. O segundo tratamento que mais ocorreu nos estudos analisados foi o Redimensionamento das imagens com cerca de 21% dos trabalhos fizeram menção ao uso dessa técnica. A fim de encontrar as técnicas de *data augmentation* utilizadas nos trabalhos selecionados foi feita uma sumarização das atividades feitas para processamento de imagem relacionadas a técnicas de *data augmentation* (Figura 4(b)). Esse detalhamento foi feito por conta que *data augmentation* foi a técnica mais citada nos trabalhos selecionados. Desta forma, a Figura 4(b) apresenta uma sumarização das técnicas de *data augmentation* com o número de ocorrências das mesmas. Como pode ser observado a técnica *flip* horizontal e Rotação aparecem indicadas em cinco trabalhos, seguido de

translação com quatro trabalhos.



(a) Técnicas de processamento

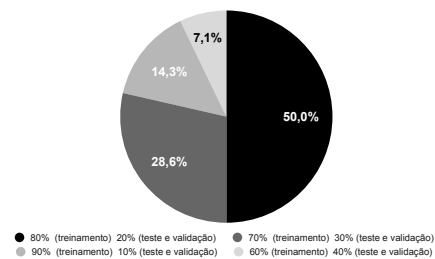


(b) Tipos de data augmentation

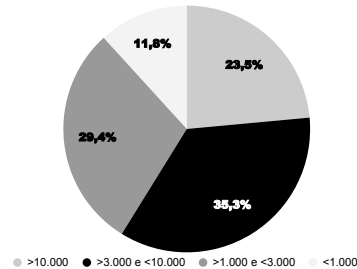
Fig. 4. Técnicas de processamento

Além disso, foi observada a estratégia adotada nos trabalhos para escolher a porcentagem de imagens para teste, treinamento e validação. A Figura 5(a) apresenta uma sumarização dos dados dos 16 estudos (um trabalho não mencionou qual foi a divisão de dados utilizada). Como pode ser observado na Figura 5(a) cerca de 50% dos trabalhos utilizam a divisão da base de dados em 80% das imagens para treinamento e 20% restante são para teste e validação. Ainda, na Figura 5(b) definiu-se um agrupamento da quantidade de imagens utilizadas nos estudos. Este agrupamento foi composto de quatro grupos: estudos que apresentaram uma quantidade de imagens maior que 10.000 imagens (cerca de 23% dos estudos), estudos em que o número de imagens foi entre 3.000 e 10.000 (cerca de 35% dos estudos), estudos que apresentaram entre 1.000 e 3.000 mil imagens (cerca de 29% dos estudos). Finalmente, grupos que apresentaram menos que 1.000 imagens (cerca de 11% dos estudos). A Figura 5(b) apresenta o quantitativo de estudos em cada grupo. Como pode ser observado o grupo entre 3.000 e 10.000 imagens foi o grupo mais representativo, com cerca de 35% dos estudos.

Ainda, a Figura 6 apresenta uma sumarização dos trabalhos em relação, o modelo utilizado (classificação ou regressão), (Figura 6(a)) se o trabalho utilizou *transfer learn-*



(a) Divisão de dados



(b) Imagens

Fig. 5. Divisão dos dados e quantidade de imagens

ing (Figura 6(b)) e se o trabalho disponibiliza a base de dados utilizada no estudo (Figura 6(c)). Finalmente, os resultados desta seção são importantes pois além de caracterizar e sumarizar os trabalhos selecionados, também podem apoiar a construção de protocolos para futuros estudos.

B. Quais são as principais RNP utilizadas?

A Figura 7 apresenta as 25 redes neurais identificadas no presente estudo e a quantidade de ocorrências dessas redes. Além disso, foram identificadas as redes utilizadas para tarefas de regressão e classificação. Como pode ser observado a rede *VGG-116* foi a rede mais utilizada para regressão com três ocorrências e a rede *SOS-Net* para classificação, com quatro ocorrências. Este resultado pode apoiar pesquisadores que desejarem escolher redes artificiais para compor seus estudos.

C. Quais são as principais características das RNP?

Nos trabalhos analisados, observou-se que as principais métricas reportadas incluem acurácia (todos os trabalhos), *recall* (75% dos trabalhos), precisão (50% dos trabalhos), MAE (*Mean Absolute Error*) (todos os trabalhos), MSE (*Mean Squared Error*) (55%) e RMSE (*Root Mean Squared Error*) (55%). Essas métricas permitem a avaliação do desempenho dos modelos, sendo essencial a escolha adequada de cada uma, especialmente considerando a diferença entre tarefas de classificação e regressão. A seleção da métrica impacta diretamente a interpretação dos resultados e a tomada de decisões sobre quais modelos apresentam melhor desempenho em diferentes contextos.

Como as métricas de acurácia e MAE foram as mais utilizadas nos trabalhos, a Figura 8 apresenta os valores encontrados para estas métricas. A Figura 8(a) apresenta a sumarização dos melhores resultados de acurácia para RNP

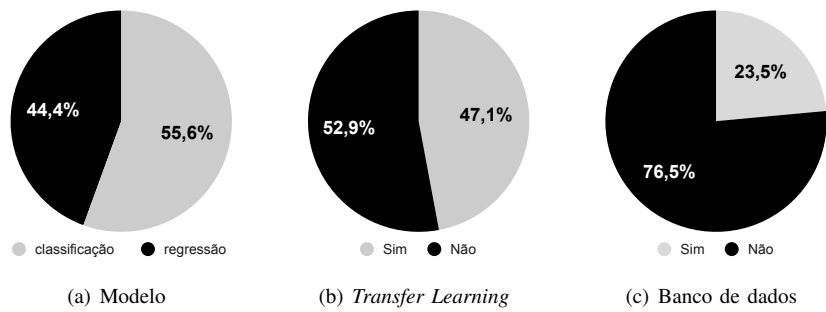


Fig. 6. Regressão/classificação, Transfer Learning e base de dados

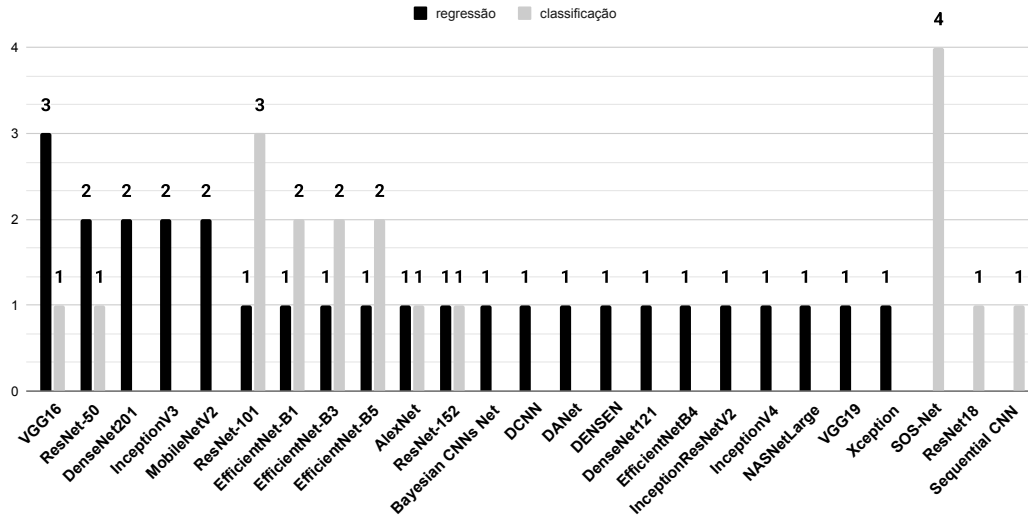


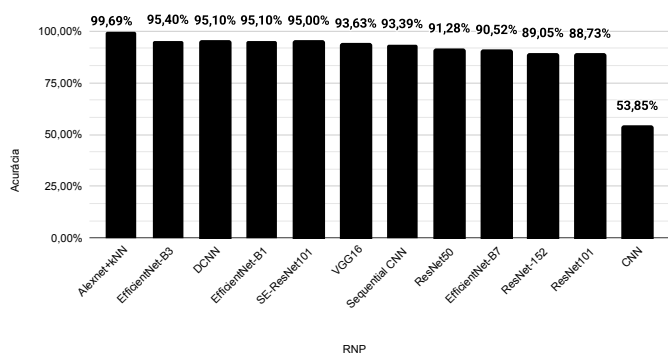
Fig. 7. Redes Identificadas

aplicadas à classificação, segundo a menção dos trabalhos selecionados. Uma vez que todos os estudos analisados reportaram a acurácia, optou-se por sumarizar os seus resultados. A acurácia é uma métrica comum e amplamente utilizada por pesquisadores da área para avaliar o desempenho de seus modelos e compará-los com resultados disponíveis na literatura. A rede neural com melhor ganho entre as demais redes analisadas foi *Alexnet+kNN* (com cerca de 99% de acurácia), conforme apresentado pela Figura 8(a). Além disso, a Figura 8(b) apresenta os melhores resultados da métrica MAE para os estudos de regressão. O MAE mede a média das diferenças absolutas entre os valores reais e os valores preditos pelo modelo. A escolha dessa métrica se justifica porque, ao contrário do MSE (Erro Quadrático Médio), que penaliza erros maiores de forma quadrática, o MAE expressa o erro médio em termos absolutos e na mesma unidade da variável predita. Essa característica facilita a análise comparativa entre diferentes abordagens e evita que *outliers* influenciem desproporcionalmente os resultados. Além disso, é uma métrica intuitiva. Conforme apresentado na Figura 8(b) a rede neural com melhor resultado foi a *SOS-NET* com o valor para a métrica MAE com cerca de 0,470 anos.

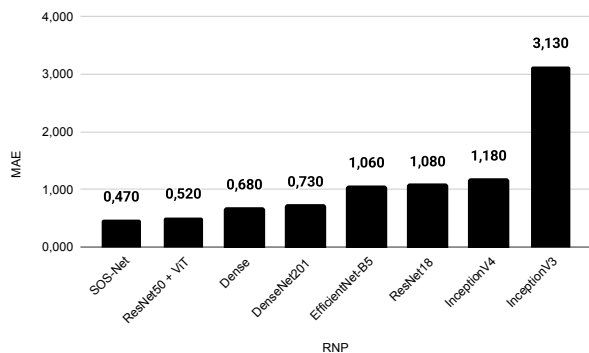
A Figura 9 apresenta o comportamento das redes neurais en-

contradas nos estudos selecionados, levando em conta a faixa etária. A Figura 9(a) apresenta, para as redes de classificação, os valores de acurácia. Já a Figura 9(b) apresenta os valores da métrica MAE para as redes de regressão. Como pode ser observado na Figura 9(a) as medianas dos três grupos de faixa etária estão muito próximas, ainda é possível observar que para o de 3 a 14 anos os dados de acurácia têm distribuição simétrica. Para o grupo acima de 25 anos, houve maior dispersão dos valores de acurácia. Dessa forma, foi observado que, para as crianças (de 3 a 14 anos), as redes apresentaram melhores resultados em comparação às demais faixas etárias. Entretanto, foi percebido que para as idades acima de 25 anos os modelos apresentam uma queda nos valores das métricas de avaliação.

Como pode ser observado na Figura 9(b) existe uma variação considerável da mediana para os grupos. Os quartis para o grupo de 3 a 14 anos são menores que os quartis do grupo de 15 a 25 que também são menores que o grupo maior de 25 anos. Ainda, há um aumento significativo do erro em termos da mediana, por exemplo: de 3 a 14 anos (o erro foi de cerca de um ano), já para o grupo de 15 a 25 anos, o erro foi de cerca de 2,5 anos e para o grupo acima dos 25 anos o erro foi de cerca de 4 anos. Desta forma, os menores erros



(a) Classificação



(b) Regressão

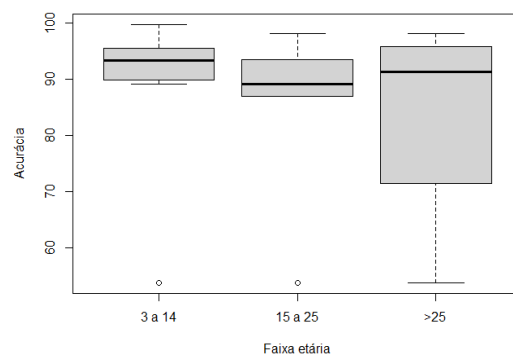
Fig. 8. Comportamento das redes neurais

ocorrem na faixa etária de 3 a 14 anos. O MAE apresenta baixa variabilidade, indicando um bom desempenho dos modelos nessa faixa e previsões mais precisas. No grupo de 15 a 25 anos, o erro médio é maior e há maior dispersão dos valores. Isso sugere que os modelos encontram mais dificuldades para prever a idade corretamente nesse intervalo. Para indivíduos acima de 25 anos, o MAE é ainda mais alto, com mediana elevada e maior variabilidade. As previsões tornam-se menos precisas e os erros são mais dispersos. Esse resultado pode estar relacionado à maior regularidade do desenvolvimento dentário em crianças, facilitando a estimativa da idade. A dificuldade de diferenciação entre 15 e 25 anos decorre da estabilidade da dentição, reduzindo sinais claros de variação etária. Acima dos 25 anos, fatores como desgaste dentário individual, implantes e uso de próteses introduzem maior variabilidade, tornando a estimativa mais complexa.

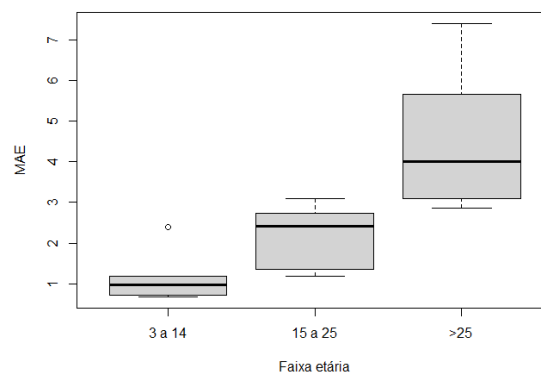
V. AMEAÇAS À VALIDADE

Foi realizado uma RSL resumindo os principais achados sobre os trabalhos que utilizam REIP, conforme apresentado nas seções anteriores. No entanto, apesar do esforço dos autores para a condução deste trabalho existem algumas ameaças que podem ter afetado a validade dos dados encontrados. Nesta seção são discutidas algumas ameaças à validade que podem afetar os achados dessa pesquisa.

Replicabilidade do estudo: O estudo foi projetado para ser aplicado nas diferentes bases de dados sem realizar alterações nos padrões de buscas, sendo necessário se adequar ape-



(a) Acurácia



(b) MAE

Fig. 9. Comportamento das redes neurais

nas às normas de utilização da biblioteca, a fim de evitar a descaracterização do estudo. Para mitigar os efeitos das discrepâncias entre as bases de dados, foi feita a definição das palavras-chave para compor a *string* de busca adaptada para cada base de dados.

Resultados alcançados: Foi realizada a sumarização das redes neurais encontradas, disponibilizadas na Seção III. No entanto, esse procedimento pode ser afetado pela forma como as redes foram agrupadas. Para minimizar esse problema, somente foram documentadas as informações que estavam explícitas nos artigos utilizados. As descobertas deste trabalho, também podem ser afetadas pelas informações disponíveis nas bases de dados, podendo não indexar todos os artigos relevantes para o foco do estudo. Para minimizar esse problema, foram selecionados seis bases de dados diferentes, como discutido na Seção II.

Generalização do estudo: Alguns fatores podem impedir a generalização dos dados encontrados na pesquisa. Por exemplo, as 25 redes neurais sumarizadas neste estudo, podem não representar adequadamente todas as redes neurais relevantes para a estimativa de idades por meio de imagens de radiografia panorâmica. As buscas realizadas neste estudo foram restritas ao período entre 2020 e 2024, podendo essa limitação afetar os resultados alcançados, uma vez que redes neurais

desenvolvidas em outro período. No entanto, este estudo foi realizado com o objetivo de coletar informações recentes sobre novas redes neurais para estimativa de idade por meio de radiografia panorâmica. A *string* de busca utilizada para identificação primária dos artigos nas bases de dados utilizadas pode ter afetado os trabalhos selecionados para compor o estudo, devido ao fato de que um dado artigo não encontrado apresente uso de redes neurais para o propósito definido neste trabalho porém não tenha sido encontrado por meio da *string* de busca utilizada. Esse problema, pode ter ocorrido devido a alguma omissão neste artigo na informações de seus metadados (título, palavras-chave e resumo). Para mitigar esse problema foi verificado que o conjunto de artigos previamente selecionado na busca *ad hoc* poderiam ser encontrados por meio da *string* de busca performada. Outra limitação diz respeito à data em que a pesquisa foi realizada, pois um artigo pode ter sido indexado após a coleta de dados utilizada.

VI. CONCLUSÕES E TRABALHOS FUTUROS

Como resultado desta RSL, foram identificadas 25 redes neurais. Dentre estas, as que apresentaram melhor desempenho foram a *AlexNet+kNN*, para classificação, e a *SOS-Net*, para regressão. Também foi observado que, para o grupo de crianças (de 3 a 14 anos), as redes atingiram melhores resultados em comparação às demais faixas etárias. Além disso, as métricas acurácia e MAE foram as mais utilizadas para avaliar o desempenho das redes neurais. Portanto, conclui-se que, embora redes neurais profundas venham sendo aplicadas com resultados promissores na classificação de imagens radiográficas, ainda há desafios significativos quanto à precisão para grupos etários acima dos 25 anos de idade. Como trabalho futuro, sugere-se a realização de um estudo empírico utilizando as redes identificadas neste trabalho, de modo a comparar sua eficácia com base em um conjunto padronizado de métricas. Tal abordagem permitirá uma análise mais precisa do ganho proporcionado pelo uso de redes neurais profundas na estimativa de idade por radiografias panorâmicas (EIROP) em diferentes grupos populacionais.

REFERENCES

- [1] N. Mualla, E. H. Houssein, and M. Hassan, "Dental age estimation based on x-ray images." *Computers, Materials & Continua*, vol. 62, no. 2, 2020.
- [2] S. Kim, Y. Lee, Y. Noh, F. Park, and Q. Auh, "Age-group determination of living individuals using first molar images based on artificial intelligence. *sci rep* 11 (1): 1073," *doi.org/10*, vol. 1038, 2021.
- [3] İ. Ataş, C. Özdemir, M. Ataş, and Y. Doğan, "Forensic dental age estimation using modified deep learning neural network," *Balkan Journal of Electrical and Computer Engineering*, vol. 11, no. 4, pp. 298–305, 2022.
- [4] S. Sathyavathi, "Human age estimation using deep convolutional neural network based on dental images (orthopantomogram)," *IETE Journal of Research*, vol. 70, no. 2, pp. 1585–1592, 2024.
- [5] Y.-C. Guo, M. Han, Y. Chi, H. Long, D. Zhang, J. Yang, Y. Yang, T. Chen, and S. Du, "Accurate age classification using manual method and deep convolutional neural network based on orthopantomogram images," *International journal of legal medicine*, vol. 135, pp. 1589–1597, 2021.
- [6] X. Wang, Y. Liu, X. Miao, Y. Chen, X. Cao, Y. Zhang, S. Li, and Q. Zhou, "Densen: a convolutional neural network for estimating chronological ages from panoramic radiographs," *BMC bioinformatics*, vol. 23, no. Suppl 3, p. 426, 2022.
- [7] H.-G. Yeom, B.-D. Lee, W. Lee, T. Lee, and J. P. Yun, "Estimating chronological age through learning local and global features of panoramic radiographs in the korean population," *Scientific Reports*, vol. 13, no. 1, p. 21857, 2023.
- [8] E. G. Mohamed, R. P. D. Redondo, A. Koura, M. S. EL-Mofty, and M. Kayed, "Dental age estimation using deep learning: A comparative survey," *Computation*, vol. 11, no. 2, p. 18, 2023.
- [9] A. Galibourg, S. Cussat-Blanc, J. Dumoncel, N. Telmon, P. Monsarrat, and D. Maret, "Comparison of different machine learning approaches to predict dental age using demirjian's staging approach," *International Journal of Legal Medicine*, vol. 135, pp. 665–675, 2021.
- [10] W. Dong, M. You, T. He, J. Dai, Y. Tang, Y. Shi, and J. Guo, "An automatic methodology for full dentition maturity staging from opg images using deep learning," *Applied Intelligence*, vol. 53, no. 23, pp. 29 514–29 536, 2023.
- [11] Y. Shi, Z. Ye, J. Guo, Y. Tang, W. Dong, J. Dai, Y. Miao, and M. You, "Deep learning methods for fully automated dental age estimation on orthopantomograms," *Clinical Oral Investigations*, vol. 28, no. 3, p. 198, 2024.
- [12] M. B. Sivri, S. Taheri, R. G. K. Ercan, Ü. Yağcı, and Z. Golrizkhatami, "Dental age estimation: A comparative study of convolutional neural network and demirjian's method," *Journal of Forensic and Legal Medicine*, vol. 103, p. 102679, 2024.
- [13] C. M. Nolla *et al.*, "The development of permanent teeth," Ph.D. dissertation, University of Michigan Ann Arbor, 1952.
- [14] R. A. Nicodemo, L. C. d. Moraes, and E. Médici Filho, "Tabela cronológica da mineralização dos dentes permanentes entre brasileiros," *Rev Fac Odontol São José dos Campos*, vol. 3, no. 1, pp. 55–6, 1974.
- [15] S. I. Kvaal, K. M. Koltveit, I. O. Thomsen, and T. Solheim, "Age estimation of adults from dental radiographs," *Forensic science international*, vol. 74, no. 3, pp. 175–185, 1995.
- [16] S.-H. Ong, H. Kim, J.-S. Song, T. J. Shin, H.-K. Hyun, K.-T. Jang, and Y.-J. Kim, "Fully automated deep learning approach to dental development assessment in panoramic radiographs," *BMC Oral Health*, vol. 24, no. 1, p. 426, 2024.
- [17] M. F. Gomes and R. S. Parpinelli, "Segmentação de imagens de fundo de olho e identificação de padrão isnt para auxílio ao diagnóstico de glaucoma," in *Proceedings of the 2023 Congresso Brasileiro de Informática em Saúde (CBIC)*, 2023, pp. 1–8.
- [18] D. H. Hoe de Castro, D. D. Ferreira, D. Z. Rodriguez, W. S. Lacerda, F. E. d. M. Borges, and C. A. S. Silva, "Utilização de curvas principais na triagem de pacientes com tuberculose," in *Proceedings of the 2023 Congresso Brasileiro de Informática em Saúde (CBIC)*, 2023, pp. 1–7.
- [19] M. Lopes Jr., R. Barbosa, and M. Fernandes, "Aprendizagem não-supervisionada aplicada à estratificação de risco de nascimentos prematuros no brasil com recorte em dados socioeconômicos," in *Proceedings of the 2021 Congresso Brasileiro de Informática em Saúde (CBIC)*, 2021, pp. 1–8.
- [20] C. C. Mu and L. Gang, "Age estimation using panoramic radiographs by transfer learning," *Chinese Journal of Dental Research*, vol. 25, no. 2, 2022.
- [21] N. Mohammad, R. Ahmad, M. H. A. Gaus, A. Kurniawan, and M. Y. P. M. Yusof, "Accuracy of automated forensic dental age estimation lab (f-dentest lab) on large malaysian dataset," *Forensic Science International*, vol. 361, p. 112150, 2024.
- [22] W. Oliveira, M. Albuquerque Santos, C. A. P. Burgardt, M. L. Anjos Pontual, and C. Zanchettin, "Estimation of human age using machine learning on panoramic radiographs for brazilian patients," *Scientific Reports*, vol. 14, no. 1, p. 19689, 2024.
- [23] S. Keele *et al.*, "Guidelines for performing systematic literature reviews in software engineering," Technical report, ver. 2.3 ebse technical report. ebse, Tech. Rep., 2007.
- [24] S. Fabbri, C. Silva, E. Hernandez, F. Octaviano, A. Di Thommazo, and A. Belgamo, "Improvements in the start tool to better support the systematic review process," in *Proceedings of the 20th international conference on evaluation and assessment in software engineering*, 2016, pp. 1–5.