

PGCOMP - Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal da Bahia (UFBA)  
Av. Milton Santos, s/n - Ondina  
Salvador, BA, Brasil, 40170-110

<https://pgcomp.ufba.br>  
[pgcomp@ufba.br](mailto:pgcomp@ufba.br)

A odontologia forense é uma área que aplica conhecimentos científicos para analisar elementos dentários, como dentes e arcadas dentárias, com o objetivo de identificar indivíduos. Essa análise é especialmente valiosa em situações como desastres em massa, investigações criminais e casos de pessoas desaparecidas. Tradicionalmente, a determinação do sexo e da idade é feita por meio de análises morfológicas e métricas de estruturas dentárias e ósseas. No entanto, essas abordagens têm limitações, como a variabilidade dos dados e a subjetividade das análises. Com os avanços no aprendizado profundo, tornou-se possível aplicar visão computacional para analisar radiografias dentárias, utilizando técnicas de classificação e regressão. Esta dissertação propõe uma abordagem Multitask-Dynamic Weighted Loss Vision Transformer-Kolmogorov-Arnold Networks (MT-DWL ViT-KAN) que combina aprendizado multitarefa com pré-treinamento autossupervisionado utilizando Mask Autoencoders, permitindo ao modelo aprender representações latentes robustas de grandes volumes de dados não rotulados. Além disso, as redes Kolmogorov-Arnold são empregadas para decompor relações complexas entre as características dentárias e os rótulos de sexo e idade, melhorando a precisão do modelo. O metodologia denominada também incorpora uma função de perda com ponderação logarítmica dinâmica, que ajusta automaticamente os pesos das tarefas durante o treinamento, garantindo um equilíbrio otimizado entre a classificação do sexo e a estimativa da idade. Em nossos experimentos, a incorporação da estratégia de Mask Autoencoders, que utiliza a arquitetura Vision Transformer, resultou em uma melhoria significativa no desempenho do modelo em comparação com o modelo EfficientNetV2-L pré-treinado na ImageNet. Além disso, ao integrar a rede Kolmogorov-Arnold na camada final multitarefa, obtivemos os melhores resultados entre todas as configurações testadas. O MT-DWL ViT-KAN alcançou um erro absoluto médio de 3,393 anos na estimativa de idade e um F1-score de 0,9421 na classificação de sexo. Esses resultados evidenciam o potencial do modelo proposto em extrair características relevantes de imagens médicas e em realizar previsões multitarefa no cenário da odontologia forense. É importante também que as avaliações futuras considerem a diversidade nas populações de grupos étnicos e no status socioeconômico para uma análise ainda mais abrangente.

Palavras-chave: visão computacional, aprendizado profundo, odontologia forense, estimativa de idade, classificação do sexo, multitarefa, Mask Autoencoder, rede Kolmogorov-Arnold, Vision Transformer

# MT-DWL ViT-KAN: Estimação simultânea de sexo e idade a partir de radiografias panorâmicas

Julian Santana Liang

Dissertação de Mestrado

Universidade Federal da Bahia

Programa de Pós-Graduação em  
Ciência da Computação

Dezembro | 2024

MSC | XXX | 2024

MT-DWL ViT-KAN: Estimação simultânea de sexo e idade a partir de radiografias panorâmicas

Julian Santana Liang

UFBA







Universidade Federal da Bahia  
Instituto de Computação

Programa de Pós-Graduação em Ciência da Computação

**MT-DWL VIT-KAN: ESTIMAÇÃO  
SIMULTÂNEA DE SEXO E IDADE A PARTIR  
DE RADIOGRAFIAS PANORÂMICAS**

Julian Santana Liang

DISSERTAÇÃO DE MESTRADO

Salvador  
13 de dezembro de 2024



JULIAN SANTANA LIANG

**MT-DWL VIT-KAN: ESTIMAÇÃO SIMULTÂNEA DE SEXO E  
IDADE A PARTIR DE RADIOGRAFIAS PANORÂMICAS**

Esta Dissertação de Mestrado foi apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Luciano Rebouças de Oliveira

Salvador  
13 de dezembro de 2024

Sistema de Bibliotecas - UFBA

XXXX Liang, Julian Santana.

MT-DWL ViT-KAN: Estimaco simultnea de sexo e idade a partir de radiografias panormicas / Julian Santana Liang – Salvador, 2024.

51p.: il.

Orientador: Prof. Dr. Prof. Dr. Luciano Rebouas de Oliveira.

Dissertao (Mestrado) – Universidade Federal da Bahia, Instituto de Computao, 2024.

1. Estimaco da idade e sexo. 2. Aprendizado profundo. 3. Odontologia forense. I. Oliveira, Luciano Rebouas de. II. Universidade Federal da Bahia. Instituto de Computao. III Ttulo.

CDU – XXX.XX.XXX

# TERMO DE APROVAÇÃO

**JULIAN SANTANA LIANG**

## **MT-DWL VIT-KAN: ESTIMAÇÃO SIMULTÂNEA DE SEXO E IDADE A PARTIR DE RADIOGRAFIAS PANORÂMICAS**

Esta Dissertação de Mestrado foi julgada adequada à obtenção do título de Mestre em Ciência da Computação e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia.

Salvador, DIA de dezembro de 2024

---

Prof. Dr. Luciano Rebouças de Oliveira  
Universidade Federal da Bahia (UFBA)

---

Profa. Dra. Patricia Ramos Cury  
Universidade Federal da Bahia (UFBA)

---

Profa. Dr. Jefferson Fontinele da Silva  
Universidade Federal da Bahia (UFBA)



*Em memoria da minha avo Teresinha Bispo.*



## AGRADECIMENTOS

Antes de tudo, gostaria de agradecer aquele que É.

Agradeço imensamente à minha mainha Izabel Santana, a quem insistiu pra eu nascer, a meu irmão Thon Lean, por todo o amor e apoio incondicional.

A toda a minha família e meus amigos por compartilhar a vida comigo.

Gostaria de expressar minha sincera gratidão ao meu orientador Luciano Oliveira, pela sua paciência, por todos os ensinamentos e a oportunidade de fazer esse mestrado.

Agradeço ao grupo de pesquisa Intelligent Vision Research Lab, em especial ao Bernardo Silva e David Lima, pelo seu apoio, agradeço a Universidade Federal da Bahia, na figura dos funcionários, do zelador ao reitor. A todos eles, devo diretamente pela minha gota em uma vastidão de conhecimento.

A todos os educadores que participaram da minha formação, quero expressar minha profunda admiração. Pois ser professor é exercer uma das profissões mais nobres, e tenho grande respeito por aqueles que dedicam seu tempo a esse importante ofício. De transformar a ciência em uma forma de arte que busca reduzir a incerteza sobre o funcionamento do universo.

“Não sou nada.  
Nunca serei nada.  
Não posso querer ser nada.  
À parte isso, tenho em mim todos os sonhos do mundo.”  
Fernando Pessoa



*prefiro o verde dos meus areais  
prefiro o som dos meus cantos no espaço  
prefiro sentir o mundo num abraço  
prefiro esse absurdo e até dizer  
que eu prefiro o sobrenatural  
prefiro ouvir uma voz no silêncio  
sentir no peito um amor quando eu penso  
no incomparável carinho da cruz  
o absurdo maior*

—GUILHERME SCARDINI (O Absurdo)



## RESUMO

A odontologia forense é uma área que aplica conhecimentos científicos para analisar elementos dentários, como dentes e arcadas dentárias, com o objetivo de identificar indivíduos. Essa análise é especialmente valiosa em situações como desastres em massa, investigações criminais e casos de pessoas desaparecidas. Tradicionalmente, a determinação do sexo e da idade é feita por meio de análises morfológicas e métricas de estruturas dentárias e ósseas. No entanto, essas abordagens têm limitações, como a variabilidade dos dados e a subjetividade das análises. Com os avanços no aprendizado profundo, tornou-se possível aplicar visão computacional para analisar radiografias dentárias, utilizando técnicas de classificação e regressão. Esta dissertação propõe uma abordagem Multitask-Dynamic Weighted Loss Vision Transformer-Kolmogorov-Arnold Networks (MT-DWL ViT-KAN) que combina aprendizado multitarefa com pré-treinamento autossupervisionado utilizando Mask Autoencoders, permitindo ao modelo aprender representações latentes robustas de grandes volumes de dados não rotulados. Além disso, as redes Kolmogorov-Arnold são empregadas para decompor relações complexas entre as características dentárias e os rótulos de sexo e idade, melhorando a precisão do modelo. O metodologia denominada também incorpora uma função de perda com ponderação logarítmica dinâmica, que ajusta automaticamente os pesos das tarefas durante o treinamento, garantindo um equilíbrio otimizado entre a classificação do sexo e a estimativa da idade. Em nossos experimentos, a incorporação da estratégia de Mask Autoencoders, que utiliza a arquitetura Vision Transformer, resultou em uma melhoria significativa no desempenho do modelo em comparação com o modelo EfficientNetV2-L pré-treinado na ImageNet. Além disso, ao integrar a rede Kolmogorov-Arnold na camada final multitarefa, obtivemos os melhores resultados entre todas as configurações testadas. O MT-DWL ViT-KAN alcançou um erro absoluto médio de 3,393 anos na estimativa de idade e um F1-score de 0,9421 na classificação de sexo. Esses resultados evidenciam o potencial do modelo proposto em extrair características relevantes de imagens médicas e em realizar previsões multitarefa no cenário da odontologia forense. É importante também que as avaliações futuras considerem a diversidade nas populações de grupos étnicos e no status socioeconômico para uma análise ainda mais abrangente.

**Palavras-chave:** visão computacional, aprendizado profundo, odontologia forense, estimativa de idade, classificação do sexo, multitarefa, Mask Autoencoder, rede Kolmogorov-Arnold, Vision Transformers



## ABSTRACT

Forensic dentistry is a field that applies scientific knowledge to analyze dental elements, such as teeth and dental arches, with the aim of identifying individuals. This analysis is especially valuable in situations such as mass disasters, criminal investigations, and missing persons cases. Traditionally, sex and age determination is done through morphological and metric analyses of dental and bone structures. However, these approaches have limitations, such as data variability and subjectivity of the analyses. With advances in deep learning, it has become possible to apply computer vision to analyze dental radiographs, using classification and regression techniques. This dissertation proposes approach Multitask-Dynamic Weighted Loss Vision Transformer-Kolmogorov-Arnold Networks (MT-DWL ViT-KAN) that combines multitask learning with self-supervised pre-training using Mask Autoencoders, allowing the model to learn robust latent representations from large volumes of unlabeled data. In addition, Kolmogorov-Arnold networks are employed to decompose complex relationships between dental features and sex and age labels, improving model accuracy. The methodology also incorporates a dynamic logarithmic weighted loss function, which automatically adjusts task weights during training, ensuring an optimized balance between gender classification and age estimation. In our experiments, incorporating the Mask Autoencoders strategy, which uses the Vision Transformer architecture, resulted in a significant improvement in model performance compared to the EfficientNetV2-L model pre-trained on ImageNet. Furthermore, by integrating the Kolmogorov-Arnold network in the final multitask layer, we obtained the best results among all tested configurations. The MT-DWL ViT-KAN achieved a mean absolute error of 3.393 years in age estimation and an F1-score of 0.9421 in gender classification. These results highlight the potential of the proposed model in extracting relevant features from medical images and in performing multitask predictions in the forensic dentistry scenario. It is also important that future assessments consider diversity in ethnic group populations and socioeconomic status for an even more comprehensive analysis.

**Keywords:** computer vision, deep learning, forensic dentistry, age estimation, sex classification, multitasking, Mask Autoencoder, Kolmogorov-Arnold Network, Vision Transformer



# SUMÁRIO

|   |    |
|---|----|
| <b>Capítulo 1—Introdução</b>  | 1  |
| 1.1 Motivação . . . . .   | 3  |
| 1.2 Objetivos . . . . .   | 3  |
| 1.2.1 Objetivo geral . . . . .  | 3  |
| 1.2.2 Objetivo específico . . . . .   | 3  |
| 1.3 Contribuições . . . . .   | 4  |
| 1.4 Mapa de capítulos . . . . .   | 4  |
| <b>Capítulo 2—Trabalhos Relacionados</b>  | 5  |
| 2.1 Classificação de sexo . . . . .   | 5  |
| 2.2 Estimativa de idade . . . . .   | 7  |
| 2.3 Classificação sexual e estimativa de idade simultânea . . . . .                             | 13 |
| 2.4 Relação com o trabalho proposto . . . . .   | 14 |
| <b>Capítulo 3—Multitask-Dynamic Weighted Loss Vision Transformer-Kolmogorov-Arnold Networks</b> | 17 |
| 3.1 Vision Transformers . . . . .   | 17 |
| 3.2 Masked Autoencoders . . . . .   | 18 |
| 3.3 Rede Kolmogorov-Arnold . . . . .  | 20 |
| 3.4 Ponderação logarítmica dinâmica nas tarefas . . . . .                                       | 21 |
| 3.5 Visão geral do MT-DWL ViT-KAN . . . . .   | 22 |
| <b>Capítulo 4—Avaliação Experimental</b>  | 25 |
| 4.1 Materiais e métodos . . . . .   | 25 |
| 4.1.1 Conjunto de dados . . . . .   | 25 |
| 4.1.1.1 Pré-treinamento autossupervisionado . . . . .   | 25 |
| 4.1.1.2 Ajuste fino . . . . .   | 26 |
| 4.1.2 Metodologia . . . . .   | 27 |
| 4.1.2.1 Protocolo do pré-treinamento autossupervisionado com MAE . . . . .                      | 28 |
| 4.1.2.2 Protocolo de detecção e remoção de imagens inconsistentes do banco de dados . . . . .   | 28 |
| 4.1.2.3 Protocolo de avaliação do ajuste fino . . . . .   | 30 |
| 4.1.2.4 Data augmentation . . . . .   | 30 |

|   |   |           |
|---|---|-----------|
| 4.2                                     | Métricas de avaliação . . . . .                                     | 30        |
| 4.3                                     | Detalhes de implementação . . . . .                                 | 31        |
| 4.4                                     | Análise Experimentais . . . . .                                     | 32        |
| 4.4.1                                   | Análise da reconstrução das imagens . . . . .                       | 32        |
| 4.4.2                                   | Análise da convergência da função de custo no ajuste fino . . . . . | 33        |
| 4.5                                     | Resultados Experimentais . . . . .                                  | 35        |
| 4.5.1                                   | Estimativa da idade . . . . .                                       | 35        |
| 4.5.2                                   | Classificação do sexo . . . . .                                     | 38        |
| 4.5.3                                   | Estudo ablativo . . . . .   | 38        |
| <b>Capítulo 5—Discussão e Conclusão</b> |   | <b>41</b> |
| 5.1                                     | Estimativa da idade . . . . .                                       | 41        |
| 5.2                                     | Classificação do sexo . . . . .                                     | 42        |
| 5.3                                     | Pontos fortes e limitações . . . . .                                | 43        |
| 5.4                                     | Trabalhos futuros . . . . .   | 43        |
| <b>Referências Bibliográficas</b>       |   | <b>45</b> |

## LISTA DE FIGURAS

|     |   |    |
|-----|---|----|
| 2.1 | Estágios de desenvolvimento dentário delimitados no método de Demirjian. A) Calcificação de alguns pontos oclusais sem fusão das diferentes calcificações. B) Fusão dos pontos de mineralização com detecção do contorno oclusal do dente. C) Fim da formação do esmalte e início da deposição de dentina. D) Formação da coroa até o limite amelocementário. E) O comprimento da raiz é menor que a altura da coroa. F) O comprimento da raiz é igual ou maior que o da coroa. G) Termina a formação de raízes; o orifício apical permanece aberto. H) Fechamento do orifício apical. Fonte: (KERMANI; YAZDI; HAGHIGHI, 2019). . . . . | 9  |
| 2.2 | Atlas do desenvolvimento e erupção dentária humana proposta por AlQah-tani, Hector e Liversidge (2010). . . . .   | 10 |
| 2.3 | Síntese do processo geral de pesquisa para estimação da idade e sexo proposto por Vila-Blanco et al. (2020). . . . .  | 11 |
| 2.4 | Abordagem de treinamento de modelo proposto por Milošević et al. (2022) para estimação de idade. . . . .  | 12 |
| 2.5 | Uma representação visual de cada etapa realizada ao criar a estrutura multitarefa. (I) Divisão de dados e método de treinamento (validação cruzada em 5 grupos), (II) avaliação de quatro modelos EfficientNet diferentes e (III) rede baseada em MTL e otimizações de função de perda. Fonte: (PRADO et al., 2024) . . . . .   | 14 |
| 3.1 | Visão geral da arquitetura ViT. A imagem é dividida em <i>patches</i> de tamanho fixo que são linearmente incorporados e recebem os <i>embeddings</i> de posição logo depois. Após a adição do <i>positional embeddings</i> , os vetores são processados por uma sequência de blocos da ViT. Fonte: (DOSOVITSKIY et al., 2021). . . . .   | 18 |
| 3.2 | Arquitetura do MAE. Fonte: (HE et al., 2024). . . . .   | 19 |
| 3.3 | Estrutura geral dos teoremas que embasam as MLPs e KANs. Fonte: (LIU et al., 2024). . . . .   | 20 |
| 3.4 | Uma representação visual de cada etapa realizada ao criar a estrutura do <i>framework</i> proposto. (I) construção de um modelo pré-treinado utilizando radiografias odontológicas, (II) análise visual do conjunto de dados empregando o modelo pré-treinado e (III) avaliação do <i>framework</i> proposto (validação cruzada em 5 <i>folds</i> ) nos conjuntos de dados filtrados. . . . .   | 22 |

|     |   |    |
|-----|---|----|
| 3.5 | Diagrama do MT-DWL ViT-KAN <i>end-to-end</i> para estimativa multitarefa de idade e sexo a partir de radiografias panorâmicas. O <i>framework</i> utiliza a arquitetura ViT-L/16 para processar as radiografias em <i>patches</i> , extraíndo características latentes por meio do <i>encoder</i> . Essas características são então repassadas para a KAN, que atua como camada final, com uma função de perda ponderada dinâmica para otimizar o aprendizado multitarefa. . . .  | 23 |
| 4.1 | Boxplot da inferência do modelo preliminar no banco de dados para realizar a detecção e remoção de imagens inconsistentes a partir de uma análise estatística e visual das imagens. . . . .   | 28 |
| 4.2 | Exemplos de imagens removidas após a análise visual: (a) A imagem possui rótulo de 6 anos de idade, mas não apresenta dentes decíduos, indicando possível inconsistência; (b) A imagem está rotulada como 63 anos, porém contém dentes decíduos, sugerindo erro no rótulo; (c) Radiografia sem cobertura adequada da região bucomaxilofacial; (d) Radiografia da cabeça desfocada, sem evidenciar a região bucomaxilar, comprometendo sua utilidade no treinamento; . . . . .   | 29 |
| 4.3 | Exemplo da reconstrução de uma imagem da radiografia dentária: imagem original, imagem mascarada e a imagem reconstruída. O percentual de <i>patches</i> mascarados é de 75%. . . . .   | 33 |
| 4.4 | A curva de perda no ajuste fino do framework proposto nos cinco <i>folds</i> utilizados na validação cruzada. . . . .   | 34 |
| 4.5 | Evolução dos coeficientes de ponderação que se ajustam dinamicamente à importância relativa das perdas durante o ajuste fino (a) coeficiente da tarefa da classificação do sexo, (b) coeficiente da estimação da idade. . .   | 34 |
| 4.6 | Estimativas de idade média dos modelos e seus valores observados. Cada ponto é uma amostra do conjunto de dados, o eixo x representa a idade real de uma amostra, e o eixo y representa a idade que o modelo estimou para essa amostra. A linha diagonal é a linha de estimativas perfeitas. Como pode ser visto, amostras mais jovens estão mais próximas da linha diagonal, enquanto amostras mais velhas têm um grau maior de variância. Também pode ser visto que o modelo não superestima ou subestima sistematicamente a idade. . . . . | 35 |
| 4.7 | Gráfico de boxplot dos resultados do erro absoluto na estimativa de idade por faixa etária. . . . .   | 37 |
| 4.8 | Gráfico de boxplot dos resultados do erro absoluto na estimativa de idade por faixa etária e sexo. . . . .  | 37 |

## LISTA DE TABELAS

|     |  |    |
|-----|--|----|
| 4.1 | Resumo da evolução do conjunto de dados de radiografias panorâmicas utilizados nos estudos de Hougaz et al. (2023), Prado et al. (2024) e Liang, Cury e Oliveira (2023), bem como nos experimentos de ajuste fino, organizados por faixa etária. . . . . | 26 |
| 4.2 | Resumo do conjunto de dados de radiografias panorâmicas utilizados organizados por faixa etária e sexo. . . . .  | 27 |
| 4.3 | Resumo do perfil estatístico das radiografia panorâmicas por faixa etária  | 27 |
| 4.4 | Os resultados da estimativa de idade por faixa etária no banco de dados de teste. O erro absoluto médio é mostrado como $\mu$ e a mediana do erro absoluto como $y$ . . . . .  | 36 |
| 4.5 | Os resultados da classificação do sexo por faixa etária e pelo sexo no banco de dados de teste. . . . .  | 38 |
| 4.6 | Comparação das arquiteturas ViT-L/16 + KAN, ViT-L/16 e Efficient-NetV2 na classificação de sexo e estimativa de idade em modelos multita-refas, a média é mostrado como $\mu$ . . . . .  | 39 |



## LISTA DE ACRÔNIMOS

|                       |  |
|-----------------------|--|
| <b>MAE</b>            | Mask Autoencoder   |
| <b>MSE</b>            | Erro Quadrático Médio  |
| <b>BCE</b>            | Entropia Cruzada Binária   |
| <b>IQR</b>            | Intervalo Interquartil   |
| <b>KAN</b>            | Kolmogorov-Arnold Network  |
| <b>MLP</b>            | Multilayer Perceptron  |
| <b>ViT</b>            | Vision Transformer   |
| <b>CNN</b>            | Convolutional Neural Network   |
| <b>ML</b>             | Aprendizado de Máquina   |
| <b>MTL</b>            | Aprendizado Multitarefa  |
| <b>MT-DWL ViT-KAN</b> | Multitask-Dynamic Weighted Loss Vision<br>Transformer-Kolmogorov-Arnold Networks |



## Capítulo

# 1

## INTRODUÇÃO

A odontologia forense é uma especialidade que utiliza conhecimentos e técnicas científicas para analisar processos físicos, químicos e biológicos nos elementos dentários, como dentes e arcadas dentárias. Essa análise pode fornecer informações cruciais sobre o indivíduo, incluindo idade, ancestralidade, origem geográfica, sexo e até mesmo hábitos ou patologias passadas (BERMAN et al., 2013). Em contextos onde a documentação é ausente ou pouco confiável, como na identificação de refugiados desacompanhados, indivíduos em processo de adoção ou corpos em estado avançado de decomposição, a determinação do sexo biológico e a estimativa da idade, a partir de radiografias da face, tornam-se imprescindíveis para a identificação correta desses indivíduos (SCHMELING et al., 2016).

Historicamente, os métodos forenses envolviam a destruição de restos dentários para análise (SOLHEIM, 1984). No entanto, as radiografias dentárias evoluíram, permitindo a adoção de abordagens não destrutivas (ECKERT; GARLAND, 1984). Atualmente, as radiografias dentárias, especialmente as panorâmicas, são as principais ferramentas utilizadas por cientistas forenses (CULAR et al., 2017), permitindo a extração de informações essenciais, de forma não invasiva, ajudando na construção de perfis biológicos de indivíduos não identificados. Contudo, técnicas convencionais que utilizam radiografias, como o método de Demirjian e Goldstein (1976), podem ser subjetivas, atribuindo pontuações qualitativas aos estágios de maturidade dentária, o que aumenta a margem para erros, principalmente em contextos de alta demanda, como desastres. Além disso, os conjuntos de dados usados no desenvolvimento e validação desses métodos frequentemente excluem casos com ausência de dentes, tratamento endodôntico ou cáries, limitando sua aplicabilidade, resultando em técnicas que não são aplicáveis a indivíduos que apresentem essas condições.

Nos últimos anos, o uso de redes neurais profundas tem se tornado cada vez mais comum na análise de radiografias odontológicas. O uso de aprendizado profundo oferece a vantagem de analisar uma ampla gama de características dentárias simultaneamente e as relações entre essas características, o que seria inviável de ser realizado manualmente, além de reduzir o tempo de processamento, funcionar mesmo quando os dentes não estão em perfeitas condições (VILA-BLANCO et al., 2020) e fornecer resultados

consistentes e reproduzíveis. Um exemplo notável é a redução de até 96% no tempo de processamento de um banco de dados de radiografias, obtida com redes neurais convolucionais (CNNs) (HEINRICH, 2024). Essas redes têm sido aplicadas em diversas áreas da odontologia e aplicações médicas, incluindo a análise de imagens odontológicas. Entre as aplicações destacam-se técnicas para segmentação da região dentária (JADER et al., 2018; SILVA; OLIVEIRA; PITHON, 2018), Numeração de dentes permanentes e decíduos (PINHEIRO et al., 2021), detecção automatizada de biofilme dentário (ANDRADE et al., 2023), detecção e classificação de dentes individuais (TUZOFF et al., 2019), identificação de tratamentos odontológicos anteriores, como endodontia (KUO et al., 2017), e até para diagnóstico de condições mais complexas, como osteoporose (CHU et al., 2018) e tumores de mandíbula (POEDJASTOETI; SUEBNUKARN, 2018). Essa abordagem aplicada em tarefas repetitivas, permite que os especialistas concentrem-se em análises mais complexas, enquanto o modelo realiza o trabalho árduo em uma fração do tempo.

No entanto, é fundamental destacar que a utilização de aprendizado profundo na estimativa de idade e sexo em radiografias odontológicas também enfrenta desafios. A disponibilidade de bancos de dados apropriados para a estimativa de idade dentária desempenha um papel importante na evolução do campo de pesquisa em odontologia forense. A qualidade dos conjuntos de dados e a confiabilidade das informações sobre a idade dos pacientes são de grande importância para o treinamento adequado dos modelos (MOHAMED et al., 2023). Além disso, é imperativo garantir a privacidade e a segurança dos dados utilizados, em conformidade com regulamentações e diretrizes éticas. Com conjuntos de dados mais robustos, bem como modelos cada vez mais sofisticados, é viável obter estimativas de idade mais precisas, o que contribui tanto para o aprimoramento do trabalho dos profissionais da área quanto para a melhoria do atendimento aos pacientes.

Predominantemente, a maioria dos métodos de aprendizado profundo lida com as tarefas de estimativa de idade e classificação de sexo de maneira individualizada. No entanto, modelos de visão computacional projetados para essas tarefas isoladamente costumam demandar grande poder computacional. Uma alternativa eficiente é a abordagem multitarefa, que integra ambas as tarefas em um único modelo, otimizando os recursos. Essa estratégia permite que o modelo aprenda informações compartilhadas entre as duas tarefas, o que é particularmente útil em contextos onde as características dentárias e ósseas fornecem pistas tanto sobre a idade quanto o sexo do indivíduo. Além de reduzir o custo computacional, essa abordagem simplifica a implementação em cenários práticos, contribuindo significativamente para o avanço das técnicas de análise de imagens odontológicas e aumentando a eficácia das práticas forenses.

Diante do exposto, este projeto propõe uma abordagem baseada em técnicas de aprendizado profundo para a estimação multitarefa de sexo e idade de indivíduos a partir de radiografias panorâmicas dentárias. O método, denominado Multitask-Dynamic Weighted Loss Vision Transformer-Kolmogorov-Arnold Networks (MT-DWL ViT-KAN), combina aprendizado multitarefa (MTL) com uma função de perda ponderada adaptativa, a arquitetura Vision Transformer (ViT) e a rede Kolmogorov-Arnold (KAN). Esse conjunto forma um *framework* end-to-end eficiente e robusto, projetado para a análise de imagens médicas no contexto da odontologia forense.

## 1.1 MOTIVAÇÃO

A crescente necessidade por métodos confiáveis na identificação de indivíduos em cenários complexos que lida com grandes volumes de dados e prazos curtos, como desastres em massa, investigações criminais e identificação de migrantes ou refugiados. Avanços recentes no processamento de imagens médicas, impulsionados por redes neurais profundas e melhorias substanciais em tecnologias de computação, transformaram a forma como analisamos radiografias e outros exames. Nesse contexto, a utilização de técnicas avançadas de aprendizado profundo como os Masked Autoencoders (MAE), capazes de aprender representações latentes de alta qualidade a partir de dados não rotulados, oferecem uma base sólida para o pré-treinamento de modelos de visão computacional. Ao capturar padrões complexos nas radiografias panorâmicas, essas técnicas permitem a criação de modelos mais resilientes a variações na qualidade das imagens e a condições adversas, como dentes ausentes ou radiografias com ruído. Complementarmente, a integração das KANs, reconhecidas por sua eficiência na aproximação de funções complexas, com a técnica de perda ponderada dinâmica, que ajusta automaticamente os pesos das tarefas durante o treinamento, no intuito de estimar simultaneamente a idade e sexo. O modelo foi avaliada por meio de uma série de experimentos para apoiar nossas decisões sobre as vantagens de usar a abordagem proposta.

## 1.2 OBJETIVOS

### 1.2.1 Objetivo geral

O objetivo deste trabalho é propor um método totalmente automático baseado em técnicas de aprendizado profundo para estimar simultaneamente tanto a idade cronológica quanto o sexo, a partir de imagens panorâmicas dentárias.

### 1.2.2 Objetivo específico

- aplicar técnicas de pré-treinamento da arquitetura ViT, utilizando MAE em radiografias panorâmicas;
- desenvolver um modelo multitarefa utilizando ajuste fino de uma ViT pré-treinada, integrado com a KAN, para realizar simultaneamente a estimativa automática de idade e a classificação de sexo;
- implementar uma função de perda com ponderação logaritmica dinâmica para balancear as tarefas de classificação e regressão;
- avaliar o modelo proposto utilizando métricas específicas para cada tarefa, como f1-score para classificação e erro absoluto médio para regressão, comparando o desempenho obtido com outros métodos existentes na literatura;

### 1.3 CONTRIBUIÇÕES

As principais contribuições alcançadas por este trabalho são as seguintes:

1. propor o *framework* MT-DWL ViT-KAN, uma abordagem multitarefa para a estimação simultânea de sexo e idade a partir de radiografias panorâmicas dentárias, utilizando ajuste fino de um modelo ViT pré-treinado em combinação com a rede KAN;
2. implementar um método baseado no pré-treinamento da arquitetura ViT utilizando MAE, uma técnica emergente de aprendizado autossupervisionado, aplicado a radiografias panorâmicas dentárias;

E foram publicados os seguintes trabalhos:

- **J. Liang**, P. Cury and L. Oliveira, "Revisiting Age Estimation on Panoramic Dental Images,"2023 36th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Rio Grande, Brasil, 2023, pp. 205-210.
- I. Prado, D. Lima, **J. Liang**, A. Hougaz, B. Peters and L. Oliveira, "Multi-task learning based on log dynamic loss weighting for sex classification and age estimation on panoramic radiographs,"2024 19th International Conference on Computer Vision Theory and Applications (VISAPP), Roma, Itália.

### 1.4 MAPA DE CAPÍTULOS

O presente trabalho está estruturado da seguinte forma:

- Capítulo 2 - explora a literatura existente relacionada à classificação, estimação, e abordagens multitarefas no contexto da análise de idade e sexo. Serão discutidos os principais métodos, comparações entre as abordagens existentes e como esses trabalhos se relacionam com o nosso estudo. Ao final, um resumo será fornecido para destacar os pontos abordados e como eles conectam com os capítulos subsequentes.

- Capítulo 3 - apresenta as técnicas e abordagens que compõem o MT-DWL ViT-KAN proposto para a classificação do sexo e estimação da idade a partir de radiografias panorâmicas.

- Capítulo 4 - detalha os materiais e métodos empregados na avaliação experimental do MT-DWL ViT-KAN, abordando o fluxo de trabalho adotado. Serão descritas as bases de dados, as métricas de avaliação, os experimentos conduzidos para validar o desempenho do sistema e os resultados obtidos.

- Capítulo 5 - análise dos resultados obtidos, destacando as principais contribuições do estudo, as limitações encontradas, e sugestões para trabalhos futuros. A conclusão resumirá os principais achados e o impacto do trabalho no campo de estudo.

## **TRABALHOS RELACIONADOS**

### **2.1 CLASSIFICAÇÃO DE SEXO**

A determinação do sexo é um processo multifatorial que envolve a análise de uma variedade de indicadores, os quais podem ser divididos em morfométricos e não métricos. Embora esses indicadores sejam valiosos, a presença de doenças, como lepra e sífilis, que causam deformação significativa do crânio e da mandíbula, ou alterações podem comprometer a acurácia da análise (BADRAN et al., 2015). Além disso, condições como a doença de Paget e a acromegalia foram associadas ao aumento, espessamento e masculinização das características ósseas em mulheres (LOTH; HENNEBERG, 1996). O dimorfismo sexual, que se manifesta progressivamente com o desenvolvimento do indivíduo, também pode ser profundamente influenciado por fatores externos, como ambiente, status socioeconômico e condições nutricionais, o que complica ainda mais a precisão da avaliação sexual (CLAUDIO et al., 2008). Apesar dessas dificuldades, a avaliação do sexo continua sendo, junto com a estimativa da idade, um dos passos fundamentais em investigações forenses (SAINI et al., 2011).

Estudos antropológicos e anatômicos revelam que, em média, as dimensões dos ossos femininos correspondem a aproximadamente 92% das dimensões dos ossos masculinos, o que equivale a uma proporção de 92/100. Isso indica que os ossos femininos não são apenas menores, mas também apresentam menor densidade óssea e menor robustez estrutural. (KRENZER, 2006). Essas diferenças morfológicas, juntamente com outras características, reforçam o papel fundamental da análise dos dentes na diferenciação sexual. Nesse contexto, os métodos odontométricos estão entre os métodos mais comumente utilizados para a avaliação do sexo. Esses métodos se baseiam na análise nas dimensões dos dentes, como méso-distal, buco-lingual, índices dentais e medidas diagonais dos dentes para avaliar quantitativamente diferenças em tamanho e forma entre homens e mulheres (HERRERA-ESCUADERO; TORO; PARADA-SANCHEZ, 2024). No entanto, muitas características dentárias, como a morfologia da coroa, o comprimento das raízes e os padrões craniofaciais, são comuns a ambos os sexos, o que pode limitar a precisão dessas estimativas (NAGARE; CHAUDHARI R. S.AND BIRANGANE; PARKARWAR, 2018). Apesar dessas limitações, a determinação do sexo também pode ser realizada com base na análise

de um único tipo dente. A medição das dimensões de caninos, tem se mostrado altamente discriminativa para a avaliação do sexo. Da mesma forma, outras estruturas dentárias, como pré-molares, primeiros e segundos molares e incisivos superiores, também apresentam variações significativas entre os sexos (RAJENDRAN; SIVAPATHASUNDHARAM, 2012).

Um estudo precursor nesse campo, realizado por Anderson e Thompson (1973), demonstrou que a largura dos caninos mandibulares e a distância intercanina eram maiores em homens do que em mulheres. Esses indicadores permitiram uma classificação correta do sexo em 74% dos casos em uma amostra de 171 indivíduos, com idades variando de 3 a 20 anos, sendo 83 do sexo masculino e 88 do sexo feminino. Além disso, proporções dentárias foram utilizadas para aprimorar a acurácia do método, evidenciando a relevância dessas variáveis na diferenciação sexual. Em outro estudo (KARAMAN, 2006), um modelo baseado nas medidas diagonais dos caninos alcançou uma precisão de 83,3%, utilizando um conjunto de dados de 60 amostras de estudantes com idades entre 16 e 19 anos. Em complemento ao estudo de Sorenti et al. (2019) que investigou a discriminação entre masculino e feminino a partir do tamanho dos molares inferiores, observando que os molares masculinos apresentam quantidades significativamente maiores de dentina, tanto em termos absolutos quanto proporcionais, enquanto as mulheres diferiram significativamente dos homens pela maior espessura relativa do esmalte. Adicionalmente, Badran et al. (2015) explorou a flexão do ramo mandibular como um indicador morfológico do dimorfismo sexual, analisando um conjunto de dados de 419 amostras e alcançando uma precisão de 94,6% para amostras femininas e 47,6% para amostras masculinas. Diante dessa discrepância, o artigo sugere a utilização de múltiplos métodos e características para uma análise mais precisa, em detrimento da dependência exclusiva de características morfológicas.

Por outro lado, a revisão conduzida por Herrera-Escudero, Toro e Parada-Sanchez (2024) examinou 143 estudos dos mais conhecidos repositórios, como Scielo, Science Direct, PubMed e Scopus, focados em métodos para estimar o dimorfismo sexual. A análise destacou que a odontometria foi o método mais frequentemente utilizado, representando 92,3% dos estudos sobre o tema. A precisão alcançada com a odontometria variou de 51% a 95,9%, evidenciando a variabilidade dos resultados desse tipo de método. Do mesmo modo, a revisão também ressaltou as limitações da odontometria, como a influência de fatores biológicos, variações individuais na dentição, além de desafios metodológicos e técnicos que podem afetar a acurácia dos resultados. Além disso, os métodos não métricos, que se baseiam em observações qualitativas, foram considerados menos confiáveis e precisos para a estimativa do dimorfismo sexual. Esses resultados, entre outros, corroboram a pesquisa de Heng, Manica e Franco (2022), que sugere que, embora as medições manuais sejam úteis, elas podem não ser suficientes como método exclusivo para determinar o sexo de um indivíduo.

Desenvolvimentos recentes no campo do processamento de imagens têm possibilitado a criação de metodologias que automatizam, de forma mais eficaz, os métodos de estimativa do sexo. Diversos estudos revisados demonstram um crescente interesse na aplicação de aprendizado de máquina (ML). Um exemplo disso é a pesquisa realizada por Kuchler et al. (2024), alcançando um f1-score de 0,79, embora utilize ainda indicadores odontométricos,

investigaram a combinação de dimensões mandibulares e dentárias, a altura do ramo mandibular e o tamanho mesiodistal do primeiro molar inferior, para a determinação do sexo utilizando regressão logística em 108 indivíduos com idades variando de 9 a 40 anos.

No contexto do uso de CNN em radiografias odontológicas, diversos estudos têm demonstrado resultados promissores. Rajee e Mythili (2021) e Ciconelle et al. (2023) empregaram a arquitetura ResNet, um modelo CNN pré-treinado. Rajee e Mythili (2021) implementaram um método de segmentação baseado em gradiente com limiar recursivo (GBRTS) para segmentar as imagens, focando em uma faixa etária de 20 a 60 anos com cerca de 1.000 imagens. Enquanto em Ciconelle et al. (2023) analisaram 207.946 radiografias, alcançando uma precisão de 95,02%, destacando que o desempenho do algoritmo variava pela resolução da imagem, idade e sexo dos indivíduos, com melhores resultados em indivíduos com mais de 16 anos, especialmente em imagens femininas.

Por sua vez, Ilić, Vodanović e Subašić (2019) e Milošević et al. (2019) utilizaram a rede VGG16 em seus estudos. Ilić, Vodanović e Subašić (2019) alcançou uma precisão de 94,3% em um conjunto de teste com 4.155 imagens, observando que o modelo teve maior precisão para pacientes adultos, mas apresentou menor acurácia em indivíduos com mais de 80 anos. Similarmente, Milošević et al. (2019) obtiveram uma precisão de 96,87% utilizando validação cruzada com 10 *folds* em 4.000 imagens dentárias, com idades variando entre 19 e 85 anos. Ambos os estudos identificaram uma leve tendência de declínio no desempenho à medida que a idade dos pacientes aumentava, atribuindo esse fato, em parte, à escassez de amostras nas faixas etárias mais avançadas, prejudicando a capacidade do modelo de generalizar para novas amostras dessa faixa etária.

No artigo de Hougaz et al. (2023), um conjunto de 16.824 radiografias panorâmicas dentárias foi utilizado para treinar e avaliar modelos de classificação de sexo. Devido ao desequilíbrio nos dados, com 37,7% de homens e 61,3% de mulheres, a métrica f1-score foi adotada como critério principal de avaliação em vez da acurácia, que pode ser enganosa em cenários com classes desbalanceadas. O f1-score, ao integrar precisão e sensibilidade, fornece uma avaliação mais equilibrada, especialmente para a classe minoritária. Diversos modelos pré-treinados da família EfficientNet foram empregados, e o modelo EfficientNet V2-L destacou-se, alcançando um f1-score de 91,43%.

Os estudos revisados na aplicação do aprendizado de máquina demonstram desempenho com variação de 79% a 96,87% de acurácia e f1-score de 91,43% Hougaz et al. (2023), com tamanhos de amostra que variam de 108 a 207.946 imagens. A maioria desses estudos focou em amostras adultas, com idades entre 9 e 86 anos, e utilizaram diferentes arquiteturas de CNNs, como ResNet e VGG16, além de modelos da família EfficientNet. Embora os resultados indiquem uma tendência de alta precisão, especialmente em populações adultas, a variabilidade nos dados e a influência de fatores como tamanho das amostras, idade, resolução das imagens e equilíbrio de gênero indicam que mais pesquisas são necessárias para validar e generalizar esses achados em diferentes contextos.

## 2.2 ESTIMATIVA DE IDADE

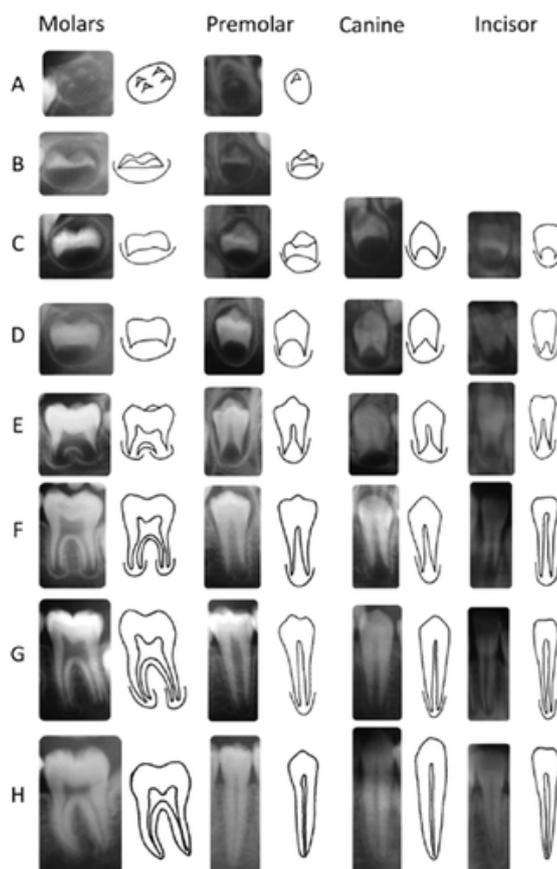
As primeiras tentativas conhecidas de utilizar dentes como indicadores de idade remontam ao trabalho de Saunders (1847), intitulado “*Teeth A Test of Age*”, que destacou a

utilidade dos dentes para essa finalidade. Desde então, a estimativa da idade a partir do desenvolvimento dental ganhou destaque, especialmente no contexto das leis trabalhistas infantis, que demandavam métodos precisos para determinar a idade das crianças. A forte correlação entre o desenvolvimento dentário e a idade cronológica, evidenciada por estudos clássicos como o de Moorrees, Fanning e Hunt (1963) e pesquisas mais recentes, como as de Cameriere, Ferrante e Cingolani (2006), consolidou o uso da odontologia como ferramenta confiável para essa estimativa. Esse método é particularmente eficaz em crianças, onde o cronograma de desenvolvimento dentário é mais previsível, permitindo uma estimativa da idade com pequena margem de erro, frequentemente medida em meses.

Para estimar a idade dentária em crianças, duas abordagens metodológicas principais são amplamente utilizadas: o método de atlas e o sistema de classificação incremental (ADSERIAS-GARRIGA et al., 2018). Os atlas, como os de Schour e Massler (1941) e *The London Atlas* de AlQahtani, Hector e Liversidge (2010) são baseados em fases de erupção dentária. O *The London Atlas* (Fig.2.2), em particular, é amplamente conhecido e foi projetado para investigar a cronologia eruptiva, categorizando indivíduos em estágios predefinidos de acordo com intervalos etários específicos. Esse atlas utiliza o desenvolvimento dentário e a erupção alveolar como indicadores para estimar a idade em indivíduos humanos. Por outro lado, os sistemas de classificação, como os propostos por Moorrees, Fanning e Hunt (1963) e Demirjian e Goldstein (1976), classificam o desenvolvimento dentário em estágios. O método de Moorrees, Fanning e Hunt (1963) estudou o desenvolvimento dentário em 14 estágios de mineralização para dentes uni e multirradiculares, determinando a idade média correspondente a cada estágio. Já o método de Demirjian e Goldstein (1976) destaca-se como um dos mais empregados na estimativa da idade (VILA-BLANCO et al., 2023a). Classificando-os em oito estágios distintos com base em características como a quantidade de depósito dentinário e mudanças na câmara pulpar (Fig.2.1).

Conforme o desenvolvimento esquelético avança, os indicadores baseados no crescimento tornam-se menos eficazes para a estimativa da idade. Após os 14 anos, o único dente ainda em desenvolvimento é o terceiro molar, embora altamente variável, permanece o indicador biológico mais confiável para essa faixa etária (LEWIS et al., 2013). No entanto, diversos métodos foram propostos para estimar a idade em jovens e adultos, baseando-se em diferentes aspectos do desenvolvimento. Um exemplo é o método de GUSTAFSON (1950), que avalia seis mudanças pós-formação dentária: atrição, periodontite, dentina secundária, deposição de cimento, reabsorção radicular e translucidez radicular. Este método pressupõe que todas as variáveis são igualmente eficazes na avaliação da idade, que o estadiamento ocorre de forma uniforme entre as seis variáveis, e que estas variáveis estatísticas são independentes (ADSERIAS-GARRIGA et al., 2018).

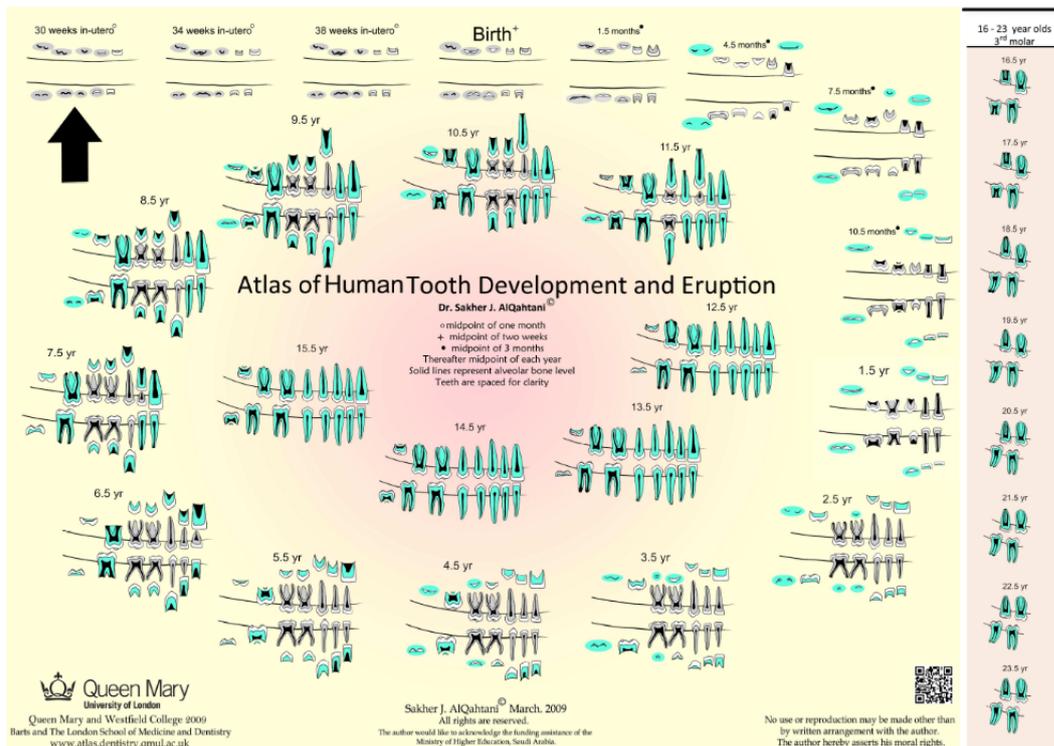
Outras abordagens incluem a análise da radiolucidez radicular e da recessão periodontal (LAMENDIN et al., 1992), assim como a medição do desenvolvimento dentário (CAMERIERE; FERRANTE; CINGOLANI, 2006). Maples (1978) introduziu um método que foca na dentina secundária e na translucidez radicular e a metodologia utilizado em Kvaal et al. (1995) que se baseia nos estágios de mineralização dos dentes. Em sua pesquisa, Cameriere, Ferrante e Cingolani (2006) mediram os ápices abertos nas raízes dos sete dentes inferiores permanentes esquerdos, normalizando essas medidas pela altura dos



**Figura 2.1** Estágios de desenvolvimento dentário delimitados no método de Demirjian. A) Calcificação de alguns pontos oclusais sem fusão das diferentes calcificações. B) Fusão dos pontos de mineralização com detecção do contorno oclusal do dente. C) Fim da formação do esmalte e início da deposição de dentina. D) Formação da coroa até o limite amelocementário. E) O comprimento da raiz é menor que a altura da coroa. F) O comprimento da raiz é igual ou maior que o da coroa. G) Termina a formação de raízes; o orifício apical permanece aberto. H) Fechamento do orifício apical. Fonte: (KERMANI; YAZDI; HAGHIGHI, 2019).

dentes, o que revelou uma correlação negativa significativa com a idade cronológica. O estudo concentrou-se na medição da distância entre os ápices abertos em dentes multirradiculares, que não possuem raízes únicas. Essas descobertas levaram ao desenvolvimento de uma fórmula de regressão que considera o sexo do indivíduo e as medidas normalizadas dos sete dentes, bem como o número total de dentes.

Enquanto o método de Kvaal et al. (1995) se baseou no fato de que a cavidade pulpar dentária se reduz com o avanço da idade devido à deposição secundária de dentina. Para estimar a idade, são calculadas as razões entre as dimensões da polpa, da raiz e do dente, o que permite a construção de um modelo preditivo de idade. No estudo, foram realizadas medições em radiografias periapicais de 100 pacientes com idades entre 20 e 87 anos, abrangendo seis tipos de dentes de cada mandíbula: incisivos centrais e laterais superiores, segundos pré-molares, incisivos laterais inferiores, caninos e primeiros pré-

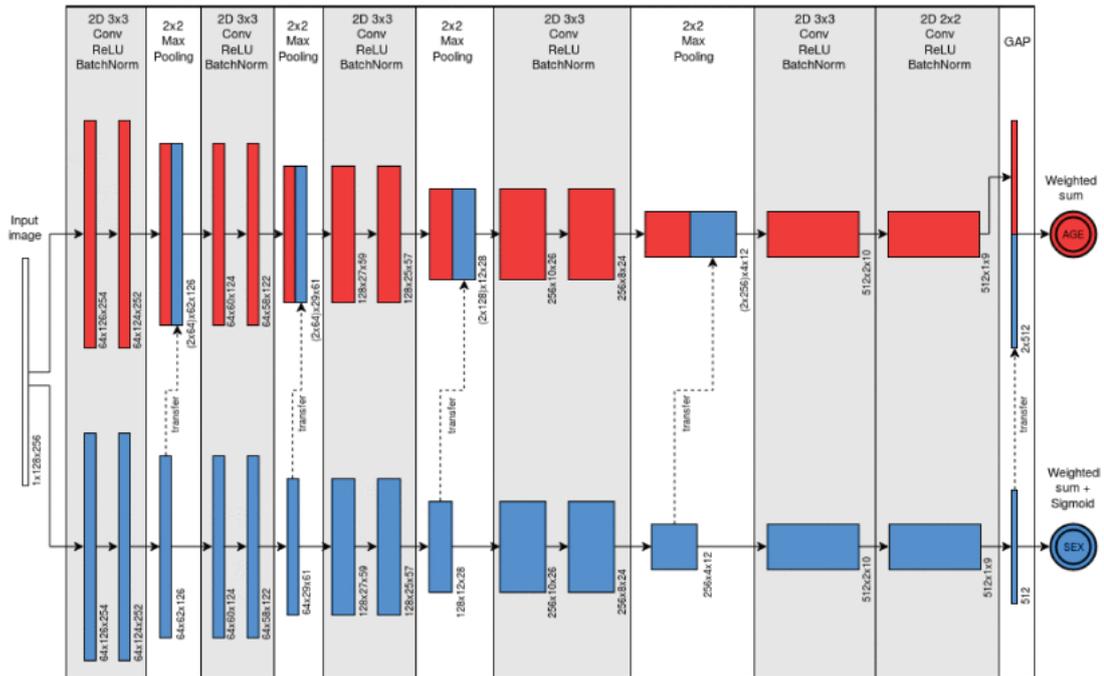


**Figura 2.2** Atlas do desenvolvimento e erupção dentária humana proposta por AlQahtani, Hector e Liversidge (2010).

molares. Entretanto, LEWIS et al. (2013) argumenta que, embora o desgaste dentário seja tradicionalmente utilizado para estimar a idade em restos antigos, sua aplicabilidade em populações modernas é limitada devido ao menor desgaste dentário observado atualmente. Além disso, fatores externos, como tratamentos ortodônticos e restaurações dentárias, também influenciam a precisão dessas estimativas, o que ressalta a importância de se utilizar técnicas que envolvem dentes não restaurados e que se baseiam na translucidez da raiz e na formação de dentina secundária para obter maior precisão.

Recentemente, diversos estudos têm tentado combinar abordagens de aprendizado profundo com a estimativa de idade na odontologia forense. Conforme estudos clássicos demonstraram, a idade e o estágio de desenvolvimento dentário podem ser estimados com precisão, e as abordagens baseadas em aprendizado profundo mantêm essa precisão ou até aprimoram essa precisão. Os métodos de estimação de idade podem ser divididos em duas categorias principais: estimativa numérica de idade e classificação por faixa etária. A maioria dos estudos, conforme destacado em uma revisão abrangente (VILA-BLANCO et al., 2023a), tem se concentrado na estimativa numérica.

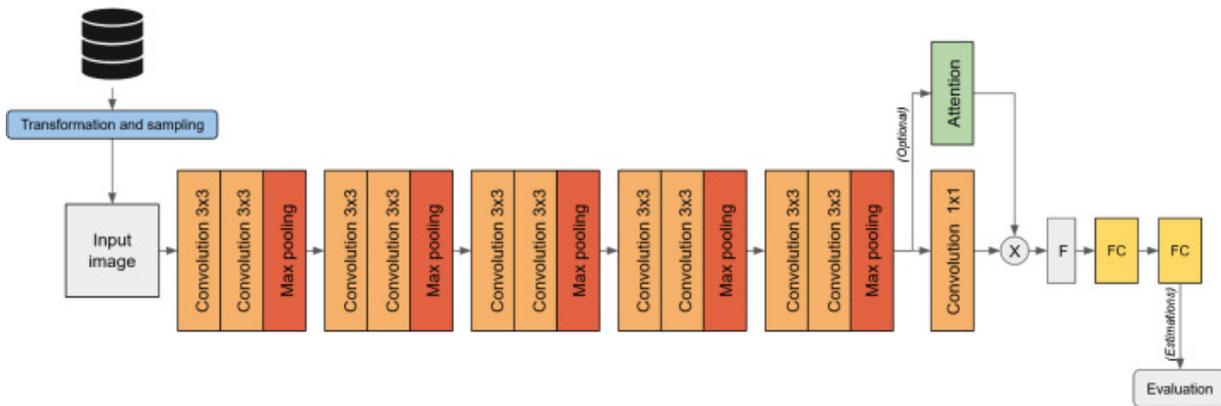
Dessa forma, o estudo de Vila-Blanco et al. (2020), que se dedica à estimativa numérica da idade a partir do aprendizado profundo, com o foco em CNNs. A CNN desenvolvida, chamada DASNet, foi projetada especificamente para essa tarefa, sem o uso de modelos pré-treinados, e incluiu a técnica compartilhamento recursos de uma segunda rede de previsão do sexo, com o objetivo de melhorar a precisão da estimativa de idade. O erro



**Figura 2.3** Síntese do processo geral de pesquisa para estimativa da idade e sexo proposto por Vila-Blanco et al. (2020).

absoluto médio foi de 2,84, porém com um desvio padrão de  $\pm 3,75$  anos. O treinamento da rede foi realizado em um conjunto de dados composto por 2.289 imagens de radiografias panorâmicas odontológicas, abrangendo um grupo populacional com média etária entre 17,1 e 25,5 anos. Em um estudo paralelo, Heinrich (2024) desenvolveram uma CNN personalizada para estimar a idade a partir de radiografias, com o objetivo de integrá-la a sistemas de identificação humana baseados em visão computacional. A CNN foi treinada em um conjunto de dados de 50.000 imagens de radiografias, abrangendo uma faixa etária de 2 a 89 anos. Os resultados indicaram um erro absoluto médio de  $3,26 \pm 3,06$  anos.

Complementando os estudos anteriores, Milošević et al. (2022) utilizaram dois tipos de imagens dentárias: 4035 imagens de radiografias panorâmicas e 76416 imagens segmentadas de dentes individuais, com idades variando entre 19 e 90 anos. O estudo avaliou o desempenho de várias CNNs pré-treinadas, como DenseNet201, InceptionResNetV2, ResNet50, VGG16, VGG19 e Xception. Além do mais, os pesquisadores conduziram experimentos com a implementação de um mecanismo de atenção, embora não tenham observado melhorias significativas nos resultados. O modelo com melhor desempenho foi o VGG16 (Fig. 2.4), com um erro de estimativa médio de 3,96 anos para imagens panorâmicas e 4,68 anos para dentes individuais. Por sua vez, Liang, Cury e Oliveira (2023) focaram na exploração de arquiteturas distintas, incluindo ConvNeXt-V2, ViT, EfficientNetV2-S, EfficientNetV2-L, EfficientNet B0 e EfficientNet B7. O melhor resultado foi alcançado pelo EfficientNet B7 com um erro absoluto médio de 4,46. Utilizando um conjunto de dados de 7.666 radiografias panorâmicas, entre as quais 4.621 são amos-



**Figura 2.4** Abordagem de treinamento de modelo proposto por Milošević et al. (2022) para estimação de idade.

tras de mulheres e 3.045 são amostras de homens de pacientes brasileiros.

De outra maneira, Yeom et al. (2023) propõem um modelo híbrido que combina ResNet50 com o ViT. Essa abordagem explora tanto características locais quanto globais das imagens para estimar a idade a partir de radiografias panorâmicas. O conjunto de dados incluiu 9663 imagens (4774 homens e 4889 mulheres, com idade média de 39 anos e 3 meses). O modelo híbrido superou os modelos individuais, com um erro absoluto médio de 2,95 anos, comparado a 3,20 anos para ResNet50 e 4,09 anos para ViT. Enquanto isso, Oliveira et al. (2024) utilizaram um conjunto de dados de 12.827 radiografias panorâmicas dentais da população brasileira, com idades entre 2,25 e 96,50 anos. Adaptando o modelo InceptionV4 e empregando técnicas de *data augmentation*, os autores alcançaram um erro absoluto médio de 3,1 anos.

Um aspecto crucial na avaliação da precisão dos modelos de estimativa de idade é a amplitude das faixas etárias e a média de idade dos conjuntos de dados utilizados. Modelos treinados em faixas etárias muito amplas podem apresentar maior representatividade de indivíduos. Por outro lado, conjuntos de dados com faixas etárias estreitas ou distribuições etárias desbalanceadas tendem a produzir modelos com desempenho limitado fora da faixa predominante. Além disso, a média de idade do banco de dados influencia diretamente a interpretação do erro médio absoluto, pois a média de idade jovem tende favorecer a avaliação do modelo. Assim, é fundamental garantir que os conjuntos de dados sejam representativos da população-alvo e que as métricas de desempenho sejam analisadas levando em conta a distribuição etária dos dados, a fim de evitar vieses que comprometam a aplicabilidade do modelo em cenários reais.

Diante disto, os estudos revisados demonstram uma ampla diversidade de abordagens baseadas em redes neurais para estimar a idade a partir de radiografias panorâmicas odontológicas. As precisões obtidas, com erro absoluto médio variando entre 2,84 e 4,68 anos, evidenciam a capacidade das CNNs, especialmente quando personalizadas ou em configurações híbridas, como ResNet50 e ViT. Os tamanhos dos bancos de dados utilizados variaram significativamente, de 4035 a 50.000 imagens, abrangendo uma faixa etária que se estende até 96,5 anos, superando o foco inicial em crianças observado nos

primeiros estudos da área. Em complemento, foram explorados fatores como o uso de mecanismos de atenção e a adaptação de arquiteturas preexistentes.

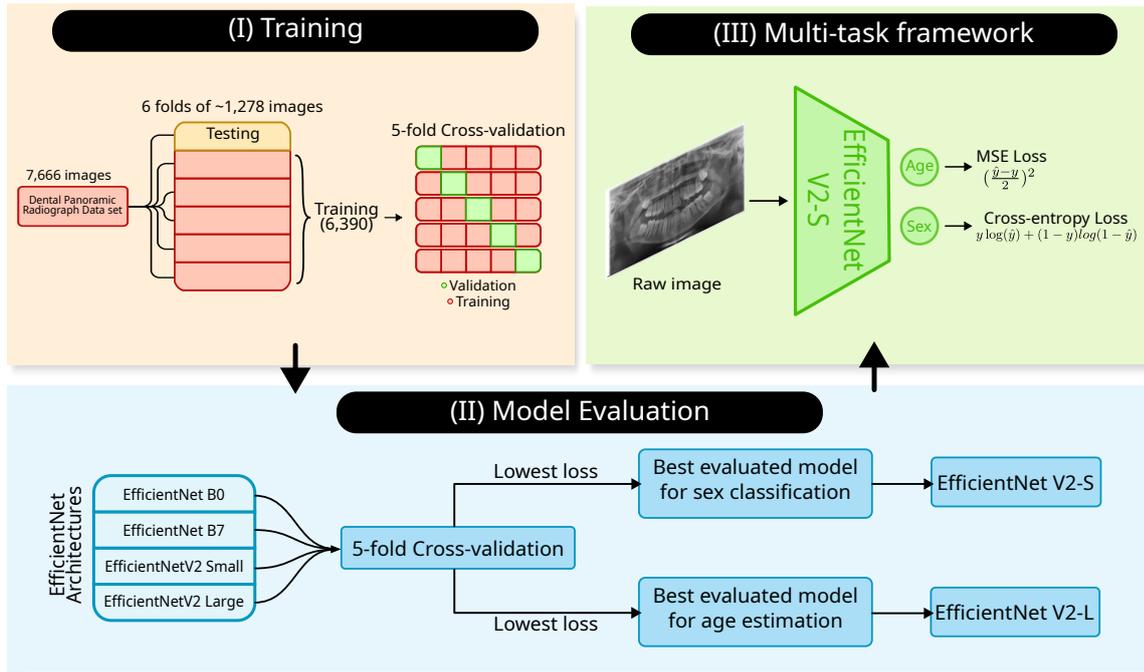
### 2.3 CLASSIFICAÇÃO SEXUAL E ESTIMATIVA DE IDADE SIMULTÂNEA

Expandindo o escopo das pesquisas anteriores, novas metodologias buscam estimar não apenas a idade ou o sexo separadamente, mas de forma integrada, ao estimar simultaneamente a idade e o sexo a partir de radiografias panorâmicas. Essa combinação de tarefas é denominada como abordagem multitarefa, essa modalidade resulta em uma economia substancial de recursos computacionais no treinamento, uma vez que um único modelo pode ser utilizado para ambas as tarefas. E ao aprender simultaneamente várias tarefas, o modelo pode beneficiar da natureza complementar das tarefas e pode capturar informações comuns e extrair características úteis que beneficiam todas as tarefas (CRAWSHAW, 2020).

Estudos como os de He et al. (2024), Hirunchavarod et al. (2024) e Milošević et al. (2022) exploraram diferentes abordagens multitarefa. He et al. (2024) analisaram 4.557 radiografias cefalométricas de crianças e adolescentes (3 a 18 anos), utilizando um modelo VGG16-MultiTask que obteve um erro absoluto médio de  $0,864 \pm 1,6$  anos para idade e um f1-score de 0,863 na classificação de sexo. Já Hirunchavarod et al. (2024) propuseram um modelo pré-treinado EfficientNetB0, aplicado a 5.132 radiografias panorâmicas de indivíduos entre 7 e 25 anos com dentes permanentes, alcançando uma precisão de 87,38% na previsão do sexo e um erro absoluto médio de 1,96 anos na estimativa da idade. De maneira similar, Milošević et al. (2022) utilizaram 86.495 imagens de dentes individuais, derivadas de 2.899 radiografias panorâmicas de indivíduos entre 19 e 86 anos, com uma proporção de 59,03% de amostras femininas. Embora tenham experimentado a abordagem multitarefa, os resultados foram inferiores aos dos modelos de tarefa única, com uma precisão de 76,41% para a previsão de sexo e um erro absoluto médio de 6,15 anos para a estimativa da idade.

Por fim, o estudo de Prado et al. (2024) implementou uma abordagem em um parâmetro adicional que se ajusta dinamicamente os pesos das funções de perda (*loss*). Na Figura 2.5, estão detalhadas a metodologia das etapas de divisão dos dados, da construção do *framework* e da avaliação da rede. No processo de avaliação, quatro modelos diferentes do EfficientNet foram testados para classificação de sexo e estimativa de idade, e a rede EfficientNetV2-L alcançou um f1-score de  $90,37\% \pm 0,54$  e um erro absoluto médio de  $5,66 \pm 0,22$ .

Esses estudos reforçam que, apesar do potencial das abordagens multitarefa, ainda há desafios a serem superados, especialmente em estudos que se concentram em populações mais jovens (abaixo de 25 anos) (HE et al., 2024; HIRUNCHAVAROD et al., 2024; VILA-BLANCO et al., 2023b). Embora modelos como VGG16-MultiTask, EfficientNetB0 e EfficientNetV2-L tenham obtido resultados positivos, com f1-score superiores a 90% e erros absolutos médios que variaram entre 0,864 e 6,15 anos, as abordagens multitarefa nem sempre superaram os modelos de tarefa única. No entanto, técnicas avançadas como a ponderação dinâmica de perdas oferecem um caminho promissor para melhorar a precisão e a eficiência computacional dessas estimativas.



**Figura 2.5** Uma representação visual de cada etapa realizada ao criar a estrutura multitarefa. (I) Divisão de dados e método de treinamento (validação cruzada em 5 grupos), (II) avaliação de quatro modelos EfficientNet diferentes e (III) rede baseada em MTL e otimizações de função de perda. Fonte: (PRADO et al., 2024)

## 2.4 RELAÇÃO COM O TRABALHO PROPOSTO

Neste capítulo, foram apresentados trabalhos relevantes na área de estimativa de sexo e idade, abordando tanto metodologias individuais quanto multitarefa, os quais serviram de base para o desenvolvimento do *framework* proposto, detalhado no próximo capítulo. O estudo de Hougaz et al. (2023), que investigou a classificação de sexo em radiografias panorâmicas, deu início à linha de pesquisa do grupo Ivisionlab na área de odontologia forense. Em seguida, um artigo de minha autoria, Liang, Cury e Oliveira (2023), revisou métodos de aprendizado profundo para a estimativa de idade a partir de radiografias panorâmicas dentárias, consolidando os fundamentos para avanços subsequentes.

No estudo mais recente de Prado et al. (2024), no qual também participei, foi implementada a primeira abordagem multitarefa para a estimação simultânea de sexo e idade. Até onde sabemos, essa é a primeira rede baseada em MTL bem-sucedida para ambas as tarefas, considerando uma ampla faixa etária. A técnica de ponderação dinâmica de perdas aplicada neste estudo visou otimizar o f1-score para a classificação de sexo e, simultaneamente, minimizar o erro médio absoluto na estimativa de idade, eliminando a necessidade de ajustes manuais exaustivos de hiperparâmetros, com pesos sendo ajustados automaticamente durante o treinamento.

Inspirados em trabalhos recentes como He et al. (2024), desenvolvemos uma arquitetura de rede neural que combina o pré-treinamento e ajuste fino baseados em MAE para

aprender representações de forma autossupervisionada e eficiente. Acrescido ao uso de KAN como uma alternativa às tradicionais Multilayer Perceptrons (MLP) nas camadas finais da rede neural.



## MULTITASK-DYNAMIC WEIGHTED LOSS VISION TRANSFORMER-KOLMOGOROV-ARNOLD NETWORKS

Neste capítulo, introduzimos os conceitos-chave que serão explorados ao longo desta pesquisa. Também contextualizamos a história das ViTs, explicando o funcionamento dos MAEs e das redes KANs. Essas técnicas são a base para o desenvolvimento deste trabalho.

### 3.1 VISION TRANSFORMERS

Na última década, as CNNs se consolidaram como a arquitetura dominante em tarefas de visão computacional, em grande parte devido à sua capacidade de capturar padrões espaciais e hierárquicos nas imagens por meio de operações de convolução em grades regulares. Essas convoluções são localmente conectadas, o que significa que cada filtro é aplicado a pequenas regiões da imagem, resultando em uma extração eficiente de características visuais, como bordas, texturas e formas simples, que são posteriormente combinadas em camadas mais profundas para detectar padrões mais complexos. Embora extremamente eficazes, as CNNs enfrentam limitações quando confrontadas com a necessidade de incorporar informações globais ou explícitas sobre a estrutura das imagens, como a integração de *tokens* de máscara ou posição dos *embeddings* posicionais, que não são naturalmente acomodados pela estrutura convolucional.

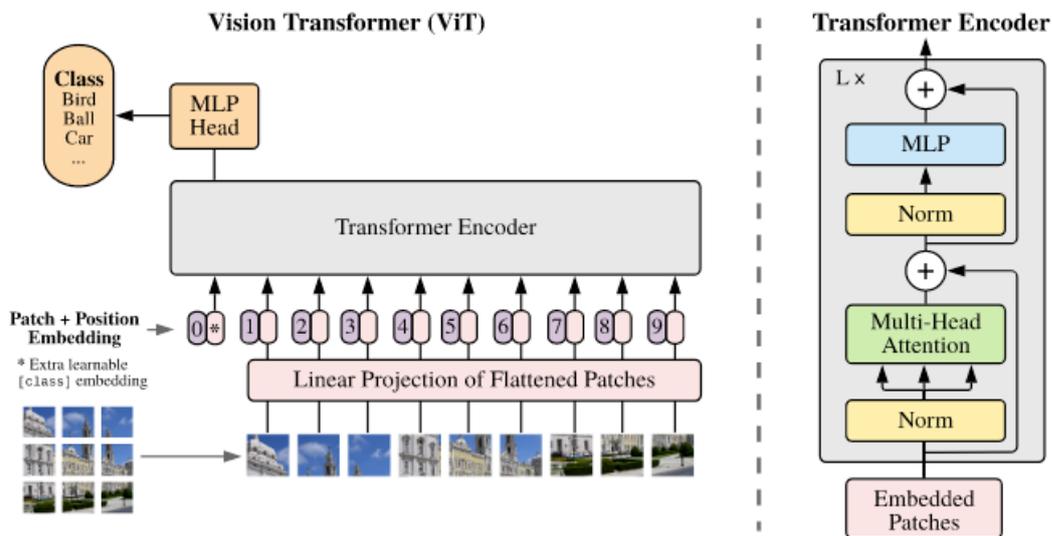
Essa limitação arquitetônica foi abordada com a introdução dos ViT (DOSOVITSKIY et al., 2021), uma adaptação dos Transformers originalmente desenvolvidos para tarefas de processamento de linguagem natural (NLP). Em vez de operar em grades regulares, como as convoluções, o ViT trata as imagens como sequências de *patches* (recortes retangulares da imagem original), onde cada *patch* é processado como um *token*, semelhante às palavras em um texto. A Figura 3.1 mostra uma visão geral da arquitetura do ViT. Essa abordagem permite que o ViT utilize mecanismos de atenção para capturar relações e contextos globais entre diferentes partes da imagem, independentemente da sua localização espacial. Além disso, os ViTs incorporam a posição dos *embeddings* diretamente

aos *tokens*, resolvendo a questão de como integrar a posição relativa dos elementos da imagem no modelo. Isso permite que os ViTs tratem não apenas a informação visual local, como as CNNs fazem, mas também a informação global de maneira mais direta e flexível, especialmente em contextos que exigem a integração de informações contextuais mais amplas, como em tarefas que envolvem imagens com estruturas espaciais complexas ou mascaradas.

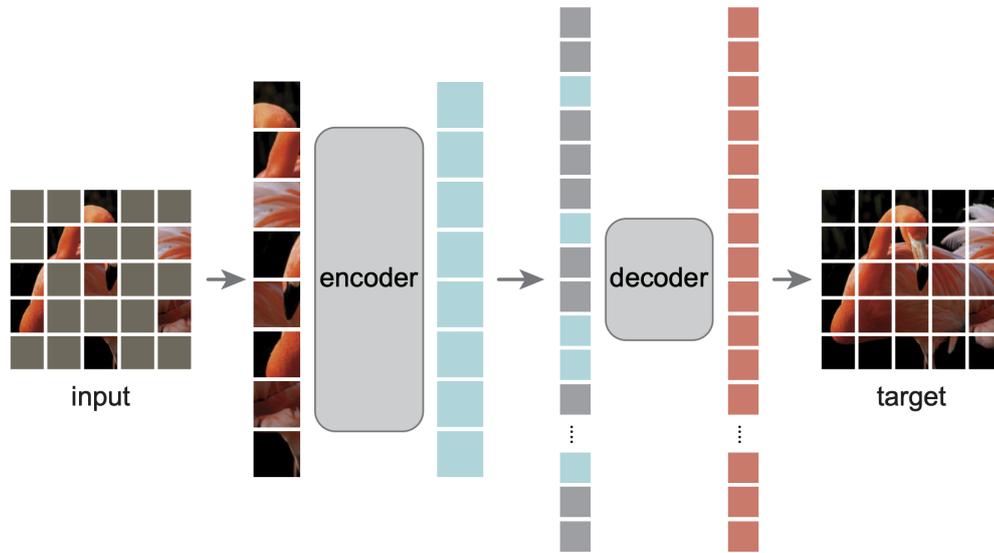
### 3.2 MASKED AUTOENCODERS

A técnica MAE (HE et al., 2024) é utilizada para pré-treinamento auto supervisionado de ViTs, inspirada por métodos de processamento de linguagem natural, onde os modelos são pré-treinados prevendo palavras ausentes em uma frase (BROWN et al., 2020). No MAE, o *pretext task* é a reconstrução da imagem original a partir de *patches* não mascarados. Ao reconstruir uma imagem onde partes substanciais foram ocultadas, o modelo é desafiado a aprender representações mais robustas e significativas dos dados de treinamento, capturando informações complexas que não são facilmente inferidas a partir dos *patches* visíveis.

O MAE opera em três etapas principais: mascaramento, codificação e decodificação. A Figura 3.2 mostra uma visão geral das etapas. Primeiramente, no mascaramento, uma parte dos *tokens* de entrada é selecionada aleatoriamente para ser ocultada. A seguir, a codificação é aplicada apenas aos *patches* visíveis, utilizando blocos de ViTs para processar esses *tokens* e produzir uma representação latente dos dados de entrada. Ao remover



**Figura 3.1** Visão geral da arquitetura ViT. A imagem é dividida em *patches* de tamanho fixo que são linearmente incorporados e recebem os *embeddings* de posição logo depois. Após a adição do *positional embeddings*, os vetores são processados por uma sequência de blocos da ViT. Fonte: (DOSOVITSKIY et al., 2021).



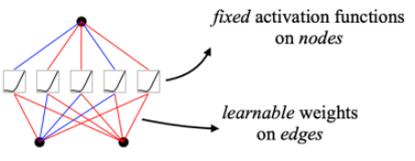
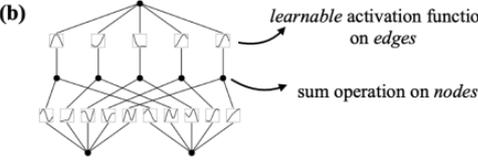
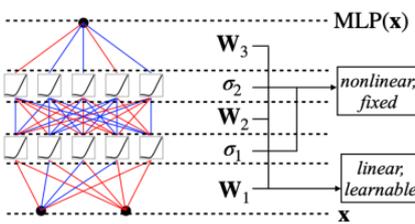
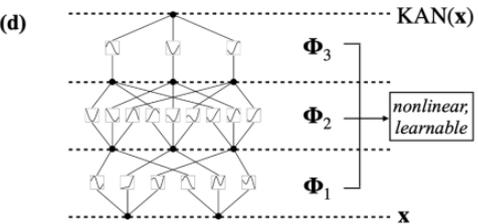
**Figura 3.2** Arquitetura do MAE. Fonte: (HE et al., 2024).

os *patches* mascarados durante a codificação, o modelo é treinado de forma eficiente, economizando memória e capacidade computacional, já que trabalha apenas com uma fração dos dados.

Após a codificação, os *tokens* mascarados são reintroduzidos para a fase de decodificação, onde o modelo tenta prever os *patches* ocultos com base no contexto fornecido pelos *patches* visíveis. O decodificador, composto por outra série de blocos ViTs, recebe tanto os patches visíveis codificados quanto os *tokens* de máscara, adicionando *embeddings* posicionais, vetores fixos e não treináveis, para incorporar informações espaciais. A tarefa do decodificador é reconstruir a imagem original a partir desse conjunto completo de *tokens*.

A função de custo no treinamento do MAE é a perda de reconstrução, calculada por meio do erro quadrático médio (MSE) entre a imagem original e a reconstruída, mas apenas nos *patches* originalmente mascarados. Isso incentiva o modelo a capturar relações globais entre os *patches* e a entender a estrutura geral da imagem. Valores normalizados de *pixels* dentro dos *patches* são utilizados como alvos de reconstrução, oferecendo melhores resultados em relação a valores brutos.

Em resumo, o MAE adota a estratégia de mascarar uma porção significativa dos *patches*, o que reduz a redundância nas imagens e cria uma tarefa desafiadora de autossupervisão que exige que o modelo desenvolva uma compreensão holística da imagem para reconstruir as partes ausentes, facilitando a aprendizagem de representações robustas. Estudos como o de Wolf et al. (2023) demonstram que os MAEs superam métodos contrastivos de autossupervisão de última geração no pré-treinamento de ViTs. O objetivo deste pré-treinamento é permitir que o modelo aprenda representações intermediárias que capturem significados semânticos úteis e complexos, fundamentais para o sucesso em

| Model             | <b>Multi-Layer Perceptron (MLP)</b>  | <b>Kolmogorov-Arnold Network (KAN)</b>   |
|-------------------|--|--|
| Theorem           | <b>Universal Approximation Theorem</b>   | <b>Kolmogorov-Arnold Representation Theorem</b>  |
| Formula (Shallow) | $f(\mathbf{x}) \approx \sum_{i=1}^{N(e)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$  | $f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left( \sum_{p=1}^n \phi_{q,p}(x_p) \right)$   |
| Model (Shallow)   | (a) <br>fixed activation functions on nodes<br>learnable weights on edges   | (b) <br>learnable activation functions on edges<br>sum operation on nodes                  |
| Formula (Deep)    | $\text{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$  | $\text{KAN}(\mathbf{x}) = (\Phi_3 \circ \Phi_2 \circ \Phi_1)(\mathbf{x})$  |
| Model (Deep)      | (c) <br>MLP(x)<br>$\mathbf{W}_3$<br>$\sigma_2$<br>$\mathbf{W}_2$<br>$\sigma_1$<br>$\mathbf{W}_1$<br>$\mathbf{x}$<br>nonlinear, fixed<br>linear, learnable | (d) <br>KAN(x)<br>$\Phi_3$<br>$\Phi_2$<br>$\Phi_1$<br>$\mathbf{x}$<br>nonlinear, learnable |

**Figura 3.3** Estrutura geral dos teoremas que embasam as MLPs e KANs. Fonte: (LIU et al., 2024).

tarefas posteriores, como a classificação de imagens.

Após o pré-treinamento, o modelo passa por um estágio de ajuste fino, no qual o decodificador é substituído por uma camada final, geralmente um MLP ou uma camada linear, que atua como classificador para emitir previsões nas tarefas subsequentes. Diferentemente do pré-treinamento, durante o ajuste fino o codificador é alimentado com o conjunto completo de *patches* de imagem, sem mascaramento, alterando o número de *patches* processados em comparação ao estágio anterior.

### 3.3 REDE KOLMOGOROV-ARNOLD

Inspiradas pelo teorema de representação de Kolmogorov-Arnold, as KANs foram propostas como uma alternativa potencial aos MLPs (LIU et al., 2024). Enquanto as MLPs utilizam funções de ativação fixas nos nós (ou "neurônios"), as KANs se distinguem ao empregar funções de ativação treináveis nas arestas (ou "pesos"). Ao invés de pesos lineares, as KANs se baseiam explicitamente no teorema de KAN para decompor funções multivariadas em funções univariadas, parametrizadas como splines, o que não ocorre nas redes neurais tradicionais. Essa diferença estrutural, embora aparentemente simples, confere às KANs vantagens notáveis em termos de precisão e interpretabilidade. A Figura 3.3 mostra algumas diferenças da MLP e da KAN.

Do ponto de vista de precisão, redes KANs significativamente menores podem alcançar resultados comparáveis ou superiores aos das MLPs, tanto teórica quanto empiricamente,

devido a leis de escala neural mais eficientes. Em termos de interpretabilidade, as KANs oferecem uma fundamentação matemática clara sobre como as entradas são transformadas e combinadas, o que potencialmente facilita a compreensão de suas operações. No entanto, essa abordagem mais estruturada pode limitar a flexibilidade em relação às redes neurais totalmente conectadas. Além disso, o processo de aprendizado dos coeficientes das funções de base nas ativações pode ser computacionalmente dispendioso, resultando em tempos de treinamento mais longos e maior uso de memória RAM, especialmente em redes mais complexas.

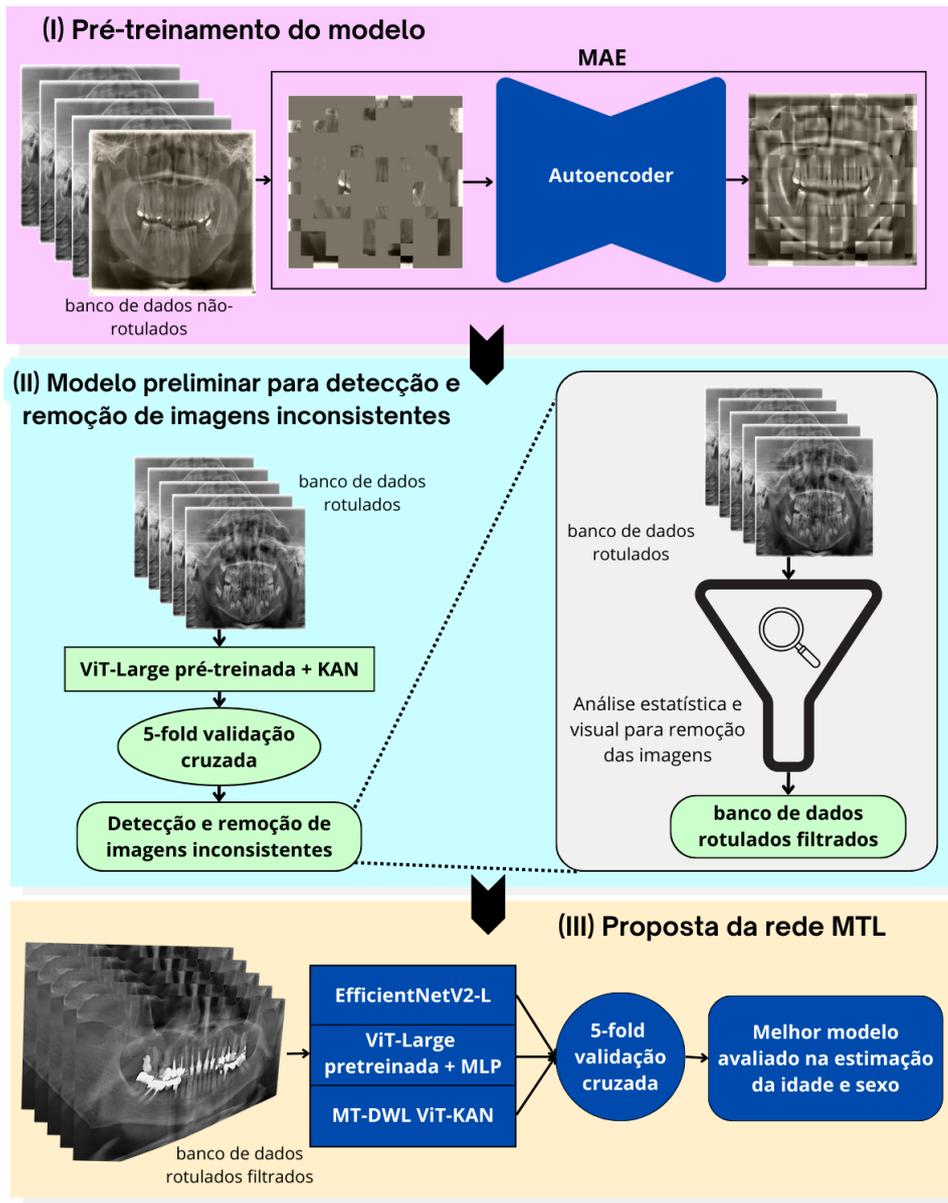
### 3.4 PONDERAÇÃO LOGARÍTMICA DINÂMICA NAS TAREFAS

No MTL, o objetivo é otimizar múltiplas funções de perda simultaneamente, necessitando de uma estratégia para combiná-las em um único valor, ou então encontrar soluções onde todas as funções de perda são otimizadas de maneira conjunta, sob a perspectiva da otimização multiobjetivo (SENER; KOLTUN, 2018). A maneira mais direta de se fazer isso é através de uma soma linear ponderada das perdas de cada tarefa individual, uma abordagem amplamente utilizada em trabalhos anteriores (CIPOLLA; GAL; KENDALL, 2017). Nessa abordagem, as perdas específicas de cada tarefa são somadas, resultando em um único valor escalar de perda. No entanto, esse método apresenta desafios, principalmente pela sensibilidade do desempenho do modelo à escolha dos pesos para cada perda. Ajustar esses hiperparâmetros pode ser caro em termos computacionais e de tempo, já que o processo de busca requer múltiplos experimentos e um ajuste manual meticuloso dos pesos, consumindo muitos dias para cada teste.

Diante desse cenário, torna-se necessário o uso de algoritmos dinâmicos capazes de ajustar automaticamente os pesos relativos das perdas com base no desempenho das tarefas, medido por métricas apropriadas. Técnicas de otimização dinâmica são essenciais em MTL, pois ajudam a equilibrar gradientes potencialmente conflitantes provenientes de diferentes tarefas, evitando a degradação do desempenho do modelo (SENER; KOLTUN, 2018). Um avanço importante nesse contexto foi a proposta de Cipolla, Gal e Kendall (2017), que introduziu uma técnica de ponderação baseada em incerteza no contexto de redes neurais convolucionais para compreensão de cena. Mais recentemente, o estudo de Prado et al. (2024) adaptou com sucesso essa abordagem para a estimação multitarefa de sexo e idade. Neste método, um parâmetro treinável adicional,  $\sigma$ , foi introduzido para modelar a incerteza dependente da tarefa, conhecida como incerteza homocedástica, que é intrínseca à natureza da tarefa e não pode ser reduzida apenas com o aumento do volume de dados de treinamento. Dessa forma, foi implementada uma ponderação logarítmica dinâmica das perdas, que permite o ajuste adaptativo dos pesos entre as diferentes tarefas de classificação e regressão, balanceando a importância de cada uma no processo de otimização. A formulação matemática dessa ponderação dinâmica pode ser expressa como:

$$L(W, \sigma_1, \sigma_2) = \frac{L_1(W, \sigma_1)}{2\sigma_1^2} + \frac{L_2(W, \sigma_2)}{\sigma_2^2} + \log(\sigma_1\sigma_2), \quad (3.1)$$

onde  $L_1$  denota a perda para a tarefa de regressão enquanto  $L_2$  é a perda para a tarefa



**Figura 3.4** Uma representação visual de cada etapa realizada ao criar a estrutura do *framework* proposto. (I) construção de um modelo pré-treinado utilizando radiografias odontológicas, (II) análise visual do conjunto de dados empregando o modelo pré-treinado e (III) avaliação do *framework* proposto (validação cruzada em 5 *folds*) nos conjuntos de dados filtrados.

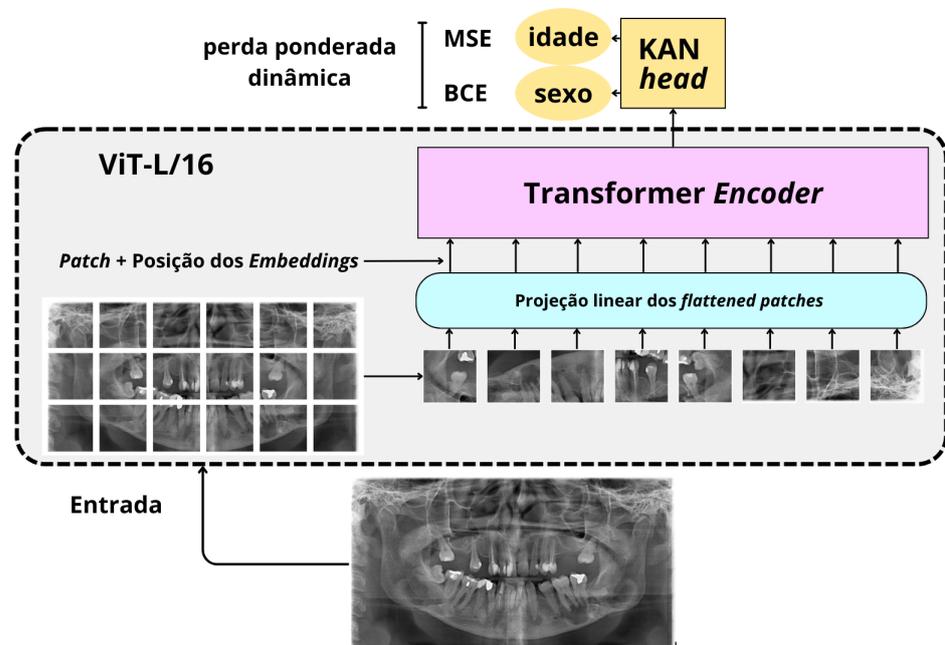
de classificação, com um conjunto de pesos  $\mathbf{W}$ .

### 3.5 VISÃO GERAL DO MT-DWL VIT-KAN

A construção do MT-DWL ViT-KAN foi organizada em três etapas principais, conforme ilustrado na Figura 3.4. Na primeira etapa, foi desenvolvido um modelo pré-treinado utilizando um MAE, aplicado a radiografias odontológicas. Esse modelo aprende re-

apresentações latentes que capturam características essenciais das imagens, mesmo na ausência de rótulos. Na segunda etapa, foi realizada uma análise visual do conjunto de dados utilizando um modelo preliminar, com o objetivo de identificar e remover inconsistências. Essas inconsistências incluíam imagens de baixa qualidade ou rótulos que não correspondiam às características do indivíduo. Essa curadoria foi para garantir que o modelo trabalhasse com dados mais consistentes, minimizando o impacto de amostras problemáticas durante o treinamento e a avaliação. Por fim, na terceira etapa, o MT-DWL ViT-KAN foi submetido por uma validação cruzada em cinco *folds*, aplicando o modelo nos dados filtrados para assegurar uma avaliação criteriosa.

O diagrama esquemático do MT-DWL ViT-KAN é ilustrado na Fig. 3.5, apresenta um fluxo de processamento *end-to-end*. O processo começa com a entrada da imagem radiográfica, seguida pela divisão em *patches* e o processamento pelo *encoder* do ViT, previamente pré-treinado utilizando um conjunto de radiografias odontológicas. A saída do ViT pré-treinado é então conectada diretamente a uma camada final, composta pela KAN, responsável por estimar simultaneamente a idade e o sexo a partir do vetor de características extraído.



**Figura 3.5** Diagrama do MT-DWL ViT-KAN *end-to-end* para estimativa multitarefa de idade e sexo a partir de radiografias panorâmicas. O *framework* utiliza a arquitetura ViT-L/16 para processar as radiografias em *patches*, extraíndo características latentes por meio do *encoder*. Essas características são então repassadas para a KAN, que atua como camada final, com uma função de perda ponderada dinâmica para otimizar o aprendizado multitarefa.



## AVALIAÇÃO EXPERIMENTAL

Neste capítulo, detalhamos os conjuntos de dados utilizados nos experimentos, as abordagens de implementação e as métricas adotadas para avaliação. Primeiramente, descrevemos as características e o processo de preparação dos dados. Em seguida, explicamos as etapas de implementação do modelo e os parâmetros experimentais, e descrevemos a metodologia em três fases principais: (i) pré-treinamento MAE, (ii) construção de um modelo preliminar para análise visual do conjunto de dados e (iii) ajuste fino do modelo pré-treinado em abordagem multitarefa com ViT-Large, integrado à rede KAN. Então, apresentamos os procedimentos de avaliação quantitativa do modelo proposto e realizamos uma comparação com estudos da literatura relevante, destacando os ganhos e limitações das abordagens exploradas.

### 4.1 MATERIAIS E MÉTODOS

#### 4.1.1 Conjunto de dados

Dois conjuntos de dados distintos foram empregados nas etapas de pré-treinamento autossupervisionado e ajuste fino do MT-DWL ViT-KAN. A evolução do conjunto de dados de radiografias panorâmicas utilizadas ao longo dos estudos está detalhada na Tabela 4.1, evidenciando as diferentes fases de curadoria e filtragem realizadas para garantir a qualidade e consistência dos dados.

**4.1.1.1 Pré-treinamento autossupervisionado** O conjunto completo de imagens não rotuladas, utilizado no pré-treinamento, consistiu em 16.824 radiografias dentárias de pacientes brasileiros. Inicialmente, a curadoria concentrou-se na classificação de sexo, conforme descrito no trabalho de Hougaz et al. (2023). A distribuição das imagens inclui 6.341 (37,7%) de pacientes do sexo masculino e 10.483 (61,3%) do sexo feminino, refletindo a heterogeneidade do banco de dados.

**Tabela 4.1** Resumo da evolução do conjunto de dados de radiografias panorâmicas utilizados nos estudos de Hougaz et al. (2023), Prado et al. (2024) e Liang, Cury e Oliveira (2023), bem como nos experimentos de ajuste fino, organizados por faixa etária.

| Faixa etária | Hougaz et al. (2023) | Liang, Cury e Oliveira (2023) | Imagens utilizadas e Prado et al. (2024) |
|--------------|----------------------|-------------------------------|--|
| 1-10         | -                    | 712                           | 680                                      |
| 11-20        | -                    | 1378                          | 1337                                     |
| 21-30        | -                    | 1920                          | 1846                                     |
| 31-40        | -                    | 1367                          | 1324                                     |
| 41- 50       | -                    | 1029                          | 995                                      |
| 51-60        | -                    | 718                           | 699                                      |
| 61-70        | -                    | 363                           | 348                                      |
| 71-80        | -                    | 132                           | 126                                      |
| 81-91        | -                    | 47                            | 42                                       |
| Total        | 16824                | 7666                          | 7397                                     |

**4.1.1.2 Ajuste fino** Em seguida, uma nova rodada de curadoria foi realizada com foco na estimativa de idade, filtrando as amostras para garantir a precisão dos rótulos de idade. Das 16.824 radiografias, apenas 7.666 continham rótulos confiáveis de idade e sexo, conforme relatado nos estudos de Prado et al. (2024) e Liang, Cury e Oliveira (2023). É importante destacar que este conjunto de dados não inclui imagens sintéticas ou aprimoramentos artificiais, com o objetivo de preservar a autenticidade e garantir uma representação fiel do mundo real. Além disso, as radiografias cobrem uma ampla gama de condições dentárias comuns na prática odontológica, como colocação de implantes dentários, cáries, periodontite, placa bacteriana, perda natural de dentes e danos à estrutura esquelética da mandíbula.

Finalmente, no intuito de aumentar a confiabilidade nos rótulos de idade, foi implementado um protocolo de detecção e remoção de imagens inconsistentes, conforme descrito na Seção 4.1.2.2. Esse protocolo teve como objetivo identificar e eliminar amostras com inconsistências ou informações discrepantes. Durante o processo de curadoria, 269 imagens foram descartadas por não atenderem aos critérios estabelecidos, como rótulos inconsistentes ou qualidade inadequada das radiografias.

Após a aplicação do protocolo, o conjunto final consistiu em 7.397 radiografias, sendo 4.480 de pacientes do sexo feminino e 2.916 do sexo masculino. Essa distribuição é apresentada em detalhes na Tabela 4.2. A faixa etária dos pacientes varia de 1 a 90 anos, com uma distribuição de 60,12% de mulheres e 39,87% de homens. Consequentemente, a distribuição da amostra por idade e sexo não é uniforme, com mais imagens de indivíduos mais jovens e uma ligeira maioria de mulheres. A idade média dos indivíduos no conjunto de dados é de 32,47 anos, o que categoriza a amostra geral como adulta, em termos de desenvolvimento dentário. Vale mencionar que as faixas etárias acima de 60 anos apresentam um número consideravelmente menor de amostras, fator que deve ser considerado na interpretação dos resultados.

**Tabela 4.2** Resumo do conjunto de dados de radiografias panorâmicas utilizados organizados por faixa etária e sexo.

| Faixa etária | Feminino | Masculino | Total |
|--------------|----------|-----------|-------|
| 1-10         | 279      | 288       | 567   |
| 11-20        | 737      | 532       | 1269  |
| 21-30        | 1175     | 685       | 1860  |
| 31-40        | 857      | 531       | 1388  |
| 41- 50       | 625      | 385       | 1010  |
| 51-60        | 465      | 265       | 730   |
| 61-70        | 231      | 153       | 384   |
| 71-80        | 82       | 59        | 141   |
| 81-91        | 29       | 19        | 48    |
| Total        | 4480     | 2917      | 7397  |

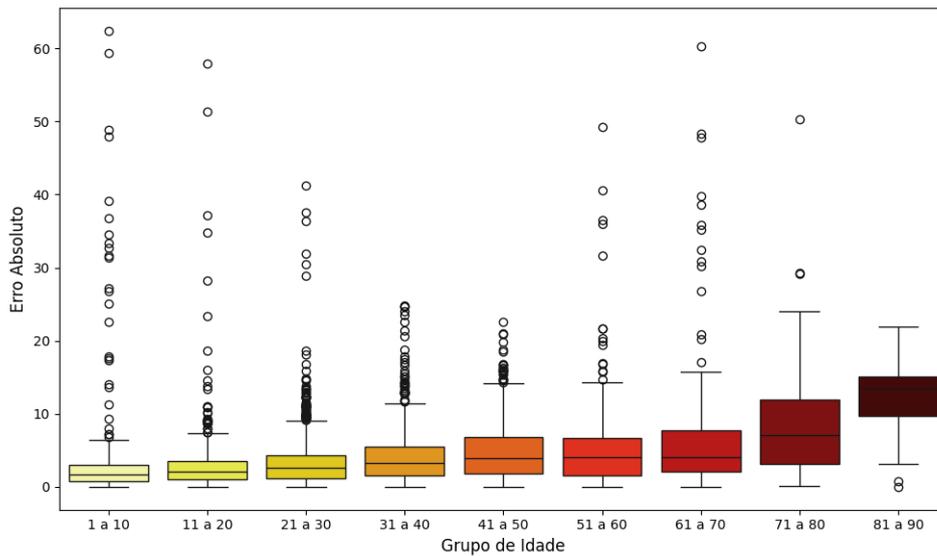
**Tabela 4.3** Resumo do perfil estatístico das radiografia panorâmicas por faixa etária

| Faixa etária | Média dos valores dos pixels | Desvio padrão médio | Moda do tamanho |
|--------------|------------------------------|---------------------|-----------------|
| 1 a 10       | 109.05                       | 57.36               | (2440, 1292)    |
| 11 a 20      | 109.71                       | 58.97               | (2440, 1292)    |
| 21 a 30      | 110.02                       | 58.98               | (2440, 1292)    |
| 31 a 40      | 110.12                       | 58.96               | (2440, 1292)    |
| 41 a 50      | 110.29                       | 59.49               | (2440, 1292)    |
| 51 a 60      | 110.19                       | 59.23               | (2440, 1292)    |
| 61 a 70      | 109.99                       | 59.09               | (2440, 1292)    |
| 71 a 80      | 110.18                       | 58.88               | (2440, 1292)    |
| 81 a 90      | 109.09                       | 59.09               | (2440, 1292)    |

As imagens foram capturadas com diferentes equipamentos, resultando em dimensões não padronizadas. A Tabela 4.3 apresenta o perfil estatístico das radiografias panorâmicas organizadas por faixa etária, incluindo a média dos valores de pixel, o desvio padrão médio e a moda do tamanho das imagens. A moda do tamanho das imagens refere-se à dimensão mais frequente em cada faixa etária específica. Em todas as faixas etárias, o tamanho de 2.440 pixels de largura por 1.292 pixels de altura foi o mais recorrente, sendo também o tamanho predominante em todo o conjunto de imagens.

#### 4.1.2 Metodologia

Este capítulo apresenta a metodologia adotada para o desenvolvimento e avaliação do método proposto. As etapas do processo metodológico estão ilustradas na Figura 3.4.

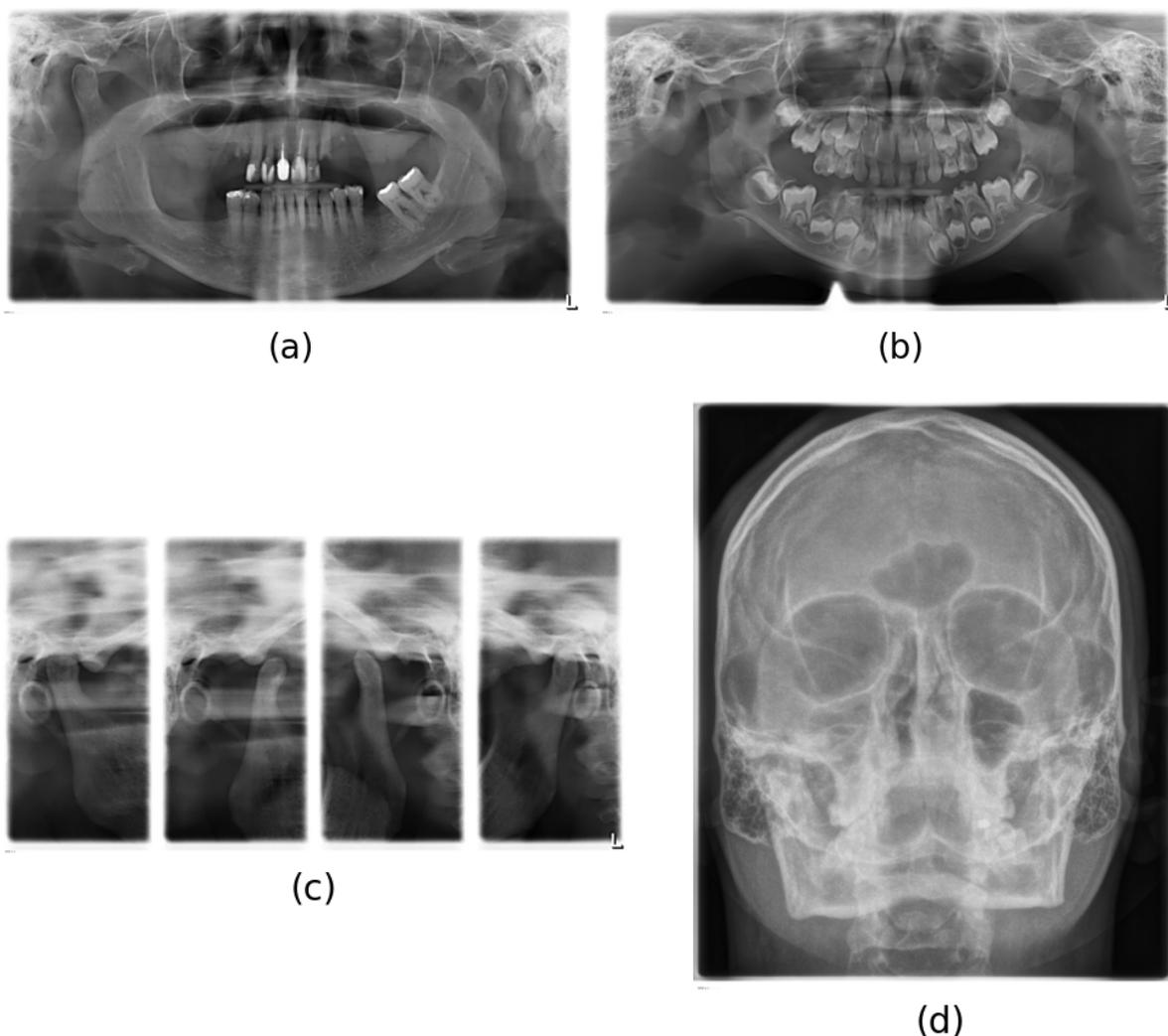


**Figura 4.1** Boxplot da inferência do modelo preliminar no banco de dados para realizar a detecção e remoção de imagens inconsistentes a partir de uma análise estatística e visual das imagens.

**4.1.2.1 Protocolo do pré-treinamento autossupervisionado com MAE** O pré-treinamento autossupervisionado foi conduzido ao longo de 800 épocas, utilizando exclusivamente imagens radiográficas, sem qualquer outra informação ou rótulo associado. Essa abordagem segue os princípios do aprendizado autossupervisionado, no qual o modelo é treinado para reconstruir partes ocultas das imagens originais. Com o MAE, uma fração significativa da imagem é mascarada aleatoriamente durante o treinamento, e o objetivo do modelo é prever essas regiões ausentes com base nas partes visíveis. E as imagens de entrada redimensionadas em  $224 \times 224$  pixels.

**4.1.2.2 Protocolo de detecção e remoção de imagens inconsistentes do banco de dados** A detecção e remoção de imagens inconsistentes foi realizada com base em uma análise estatística e visual das radiografias. Esse processo teve como objetivo identificar e eliminar imagens que continham possíveis inconsistências nos rótulos de idade dos pacientes. Imagens que apresentavam discrepâncias significativas em relação às informações fornecidas foram descartadas, evitando a introdução de ruído na fase de ajuste fino. Essa verificação auxilia em alcançar maior confiança nos rótulos, assegurando que dados mais consistentes fossem utilizados para o treinamento.

Inicialmente, um modelo preliminar foi treinado com o conjunto de dados original, composto por 7.666 imagens, utilizando validação cruzada com 5  *folds*. Após o treinamento, para cada modelo gerado, foram feitas inferências sobre todas as imagens para identificar aquelas com erros discrepantes. A análise dos resíduos foi apoiada por um boxplot dos erros médios por faixa etária, Figura 4.1, permitindo remover as imagens cujos resíduos estavam acima do limite superior, que são consideradas como  *outliers*. E foram analisadas visualmente as imagens com resíduos acima do terceiro quartil.



**Figura 4.2** Exemplos de imagens removidas após a análise visual: (a) A imagem possui rótulo de 6 anos de idade, mas não apresenta dentes decíduos, indicando possível inconsistência; (b) A imagem está rotulada como 63 anos, porém contém dentes decíduos, sugerindo erro no rótulo; (c) Radiografia sem cobertura adequada da região bucomaxilofacial; (d) Radiografia da cabeça desfocada, sem evidenciar a região bucomaxilar, comprometendo sua utilidade no treinamento;

Com base nessa análise, foram eliminadas imagens de baixa qualidade, incluindo aquelas com erros na composição, radiografias que não atendiam aos critérios de uma imagem panorâmica adequada, ou que apresentavam ruído excessivo e perda de contraste. A remoção desses dados era fundamental para evitar que o modelo aprendesse padrões irrelevantes ou ruídos, comprometendo sua capacidade de generalização. Exemplos dessas imagens eliminadas são apresentados na Figura 4.2. Assim, essa etapa de análise visual das imagens teve como objetivo aumentar a consistência e a integridade do conjunto de dados utilizado durante o treinamento

**4.1.2.3 Protocolo de avaliação do ajuste fino** Para avaliação do ajuste fino do modelo pré-treinado, foi empregada a validação cruzada com 5 *folds*, uma estratégia que reduz o viés dos resultados. As 6.373 imagens do conjunto de treino e validação, representando 83,12% do total de amostras, foram divididas em cinco subconjuntos de tamanho semelhante. A divisão foi balanceada considerando tanto o gênero quanto a faixa etária, garantindo a representatividade em cada iteração. Em cada rodada, quatro subconjuntos foram usados para treinamento, enquanto o quinto foi reservado para validação, alternando-se os subconjuntos a cada repetição. Assim, todas as imagens foram aproveitadas tanto para treinamento quanto para validação em algum momento. Essa abordagem aumenta a robustez e a capacidade de generalização do modelo, reduzindo o risco de sobreajuste a uma única divisão específica dos dados.

Após a conclusão da validação cruzada, as 1.245 imagens restantes, equivalentes a 16,88% do total, foram usadas como conjunto de teste para uma avaliação final em dados não vistos anteriormente, assegurando uma mensuração objetiva do desempenho. O resultado final foi obtido pela média das métricas de classificação e regressão dos cinco modelos treinados e descrito na Seção 4.5.

**4.1.2.4 Data augmentation** Durante o ajuste fino, foi implementada a estratégia de *data augmentation* para cada imagem, incluindo operações como redimensionamento e inversão horizontal, com o objetivo de aumentar a variabilidade e a robustez do modelo. Nesse contexto, utilizou-se a técnica RandAugment, disponível no pacote timm. Essa abordagem aplica automaticamente uma combinação de transformações, como ajustes de brilho, contraste, rotação, corte aleatório, desfoque, entre outras, com intensidades controladas. Ao introduzir essas distorções nas amostras de treino, o RandAugment contribui para melhorar a capacidade do modelo de lidar com variações visuais, reduzindo o risco de sobreajuste e promovendo um aprendizado mais generalizável.

## 4.2 MÉTRICAS DE AVALIAÇÃO

As principais métricas a serem utilizadas para a avaliação quantitativa são as comumente empregadas nas tarefas de classificação e regressão: f1-score e erro absoluto médio, respectivamente. Nesse contexto, a utilização da f1-score é relevante para a avaliação do desempenho de algoritmos de classificação, pois ela leva em consideração tanto a precisão quanto a revocação. Isso é especialmente importante quando a distribuição de classes é desigual. Quanto às métricas de avaliação para modelos de regressão, o erro absoluto médio fornece uma interpretação direta da magnitude média dos erros e é preferível em cenários onde é necessária a robustez a *outliers*.

As métricas f1-score e erro absoluto médio são representadas pelas seguintes equações:

- f1-score:

$$f1\text{-score} = \frac{2 \times TP}{2 \times TP + FP + FN}, \quad (4.1)$$

onde:

- $TP$  representa o número de verdadeiros positivos;

- $FP$  representa o número de falsos positivos;
  - $FN$  representa o número de falsos negativos.
- erro absoluto médio:

$$\text{erro absoluto médio} = \sum_{i=1}^N |\hat{y}_i - y_i| \quad (4.2)$$

onde:

- $N$  é o número total de amostras;
- $\hat{y}_i$  é o valor observado;
- $y_i$  é o valor predito.

### 4.3 DETALHES DE IMPLEMENTAÇÃO

O código foi desenvolvido utilizando o *framework* de aprendizado profundo PyTorch. O treinamento dos modelos foi realizado com um tamanho de lote de 16 em um ambiente equipado com duas GPUs: uma Nvidia GeForce RTX de 12 GB e uma Nvidia GeForce RTX 3060, também com 12 GB de memória. A configuração do sistema incluiu ainda 16 GB de memória RAM e um processador AMD Ryzen 7, oferecendo capacidade computacional adequada para o processamento eficiente dos dados.

O modelo escolhido foi o ViT-Large Patch16 (ViT-L/16), uma arquitetura baseada em ViT, que divide as imagens em blocos de  $16 \times 16$  pixels para processamento eficiente. A entrada foi configurada para  $224 \times 224$  pixels, estabelecendo o tamanho das amostras processadas por iteração. O treinamento foi realizado ao longo de 100 épocas, garantindo tempo suficiente para que o modelo convergisse, equilibrando o risco entre sobreajuste e subajuste. Para a otimização, utilizamos o AdamW, uma variante do algoritmo Adam. Diferentemente do Adam tradicional, que incorpora a regularização diretamente no gradiente, o AdamW separa o cálculo do decaimento do peso da atualização do gradiente, melhorando a convergência e a capacidade de generalização do modelo. Adotamos um *weight decay* de 0,01 como regularização para evitar sobreajuste, penalizando pesos excessivamente altos. A taxa de aprendizado inicial foi definida como 0,001, controlando a velocidade de ajuste dos parâmetros do modelo. Além disso, configuramos um mínimo de taxa de aprendizado em  $1e^{-06}$ , garantindo que a aprendizagem não cesse prematuramente, mesmo nos estágios finais do treinamento.

A função de perda variou de acordo com a tarefa: utilizamos a entropia cruzada binária (BCE) com logit para classificação e o MSE para regressão. A entropia cruzada binária com logit mede a discrepância entre a distribuição prevista e a real, combinando a função sigmoide com a entropia cruzada em uma única etapa. Essa abordagem é particularmente eficaz em tarefas com alvos binários, como a previsão do sexo a partir de radiografias, eliminando a necessidade de calcular explicitamente a sigmoide fora da função de perda. Isso melhora a estabilidade numérica, especialmente quando as probabilidades se aproximam de 0 ou 1, evitando erros de arredondamento. A BCE com logit é definida da seguinte forma:

$$\sigma(\hat{y}_i) = \frac{1}{1 + e^{-\hat{y}_i}}, \quad (4.3)$$

$$\text{Perda}_{BCE_{logit}} = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(\sigma(\hat{y}_i)) + (1 - y_i) \cdot \log(1 - \sigma(\hat{y}_i))] \quad (4.4)$$

Onde:

- $N$  é o número total de amostras,
- $y_i$  é o rótulo verdadeiro,
- $\hat{y}_i$  é a predição do modelo, e
- $\sigma(\hat{y}_i)$  é a função sigmoide aplicada à predição.

A perda em MSE tem o objetivo de minimizar a diferença entre as previsões contínuas e os valores reais. O MSE calcula o quadrado da diferença entre os valores verdadeiro e predito, o que amplifica erros maiores. Essa característica faz com que o modelo seja fortemente penalizado por previsões distantes dos valores reais, incentivando uma maior precisão geral. Essa propriedade é especialmente útil em cenários em que desvios acentuados são indesejáveis, como na análise de dados médicos e forenses, onde previsões precisas são cruciais. A função MSE é definida como:

$$\text{Perda}_{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (4.5)$$

Onde:

- $N$  é o número total de amostras;
- $\hat{y}_i$  é o valor observado;
- $y_i$  é o valor predito.

## 4.4 ANÁLISE EXPERIMENTAIS

### 4.4.1 Análise da reconstrução das imagens

A análise qualitativa da reconstrução das imagens é essencial para avaliar o desempenho do pré-treinamento com o MAE. Na Figura 4.4.1 são apresentados resultados da reconstrução de imagens pela MAE treinada com o modelo ViT-L/16, com uma amostra aleatória do conjunto de validação. A Figura é composta pela imagem original, imagem mascarada e a reconstrução realizada pelo MAE.

A radiografia completa serve como referência visual integral, apresentando todos os detalhes anatômicos que o modelo deve ser capaz de inferir, mesmo com parte das informações ocultas. Durante o pré-treinamento, 75% dos *patches* da imagem são mascarados,



**Figura 4.3** Exemplo da reconstrução de uma imagem da radiografia dentária: imagem original, imagem mascarada e a imagem reconstruída. O percentual de *patches* mascarados é de 75%.

o que cria uma tarefa desafiadora para o modelo. Essa remoção aleatória de blocos incentiva o MAE a desenvolver uma compreensão holística da estrutura das radiografias, preenchendo lacunas com base em padrões aprendidos.

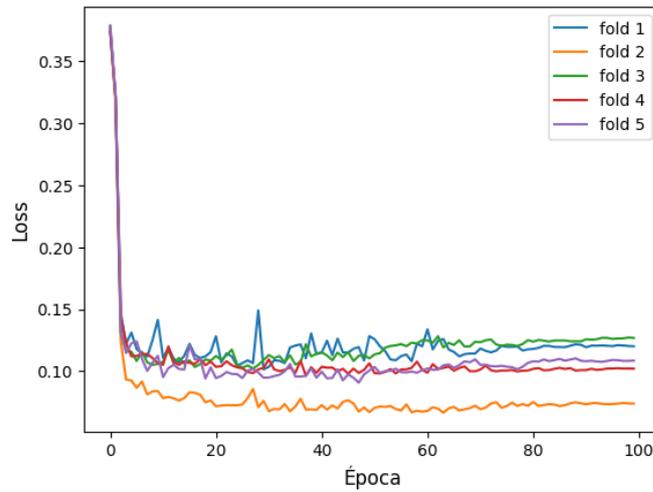
A reconstrução obtida é uma aproximação do conteúdo original, onde o modelo tenta prever os pixels ausentes. A análise qualitativa indica que o modelo preserva estruturas principais, como contornos dentários e regiões anatômicas essenciais, ainda que alguns detalhes finos e nuances de contraste não sejam perfeitamente recuperados. É importante destacar que, na imagem reconstruída, os *patches* não mascarados não são reintroduzidos, o que facilita a avaliação direta da capacidade preditiva do modelo.

Conforme observado na Figura 4.4.1, a reconstrução apresenta boa qualidade, capturando características relevantes das radiografias dentárias. Isso sugere que o modelo aprendeu representações úteis e coerentes, o que é fundamental para futuras tarefas supervisionadas. Adicionalmente, a fidelidade da reconstrução demonstra a capacidade do MAE em lidar com dados ruidosos ou parcialmente indisponíveis, tornando-o uma abordagem robusta para cenários práticos.

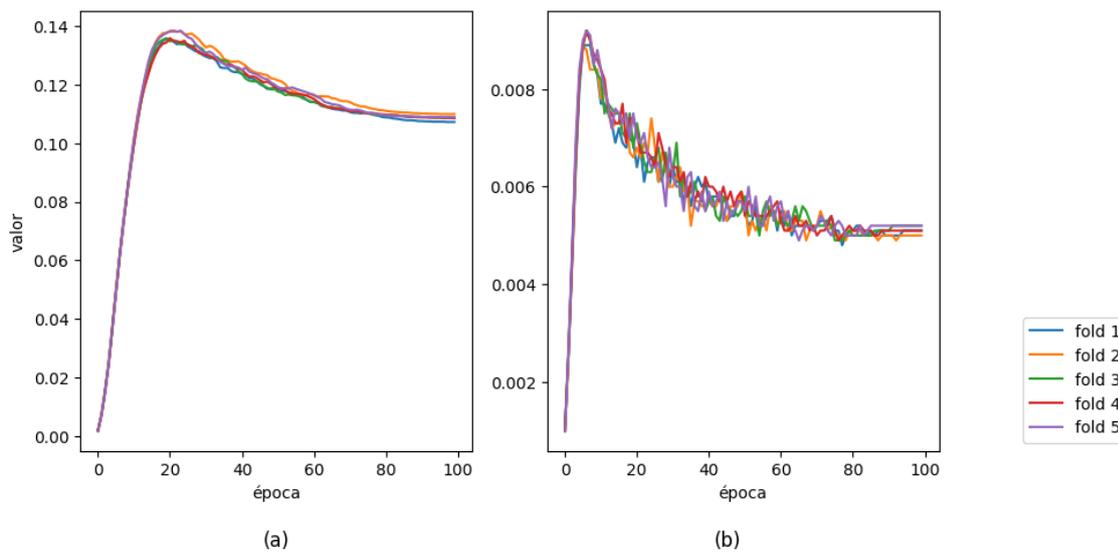
#### 4.4.2 Análise da convergência da função de custo no ajuste fino

A Figura 4.4 apresenta a evolução da perda ao longo das 100 épocas de treinamento para cada um dos cinco *folds* utilizados na validação cruzada. Embora as curvas exibam variações sutis, todas convergem para uma faixa semelhante, indicando que o modelo manteve consistência entre os diferentes subconjuntos. A estabilização das curvas nas últimas épocas sugere a ausência de sobreajuste significativo, com a perda permanecendo estável mesmo após várias iterações.

Já a Figura 4.5 ilustra a evolução dos coeficientes de ponderação durante o ajuste fino do modelo em diferentes *folds*, mostrando como o modelo ajusta dinamicamente a importância relativa das perdas. Esse processo é crucial para equilibrar as tarefas ao longo do treinamento. No início, observa-se um crescimento acentuado dos coeficientes, indicando que o modelo ainda está se adaptando e buscando um ponto de equilíbrio. Com o decorrer das épocas, esses coeficientes diminuem gradualmente e se estabilizam, sugerindo que o modelo alcançou um ajuste adequado para ambas as tarefas.

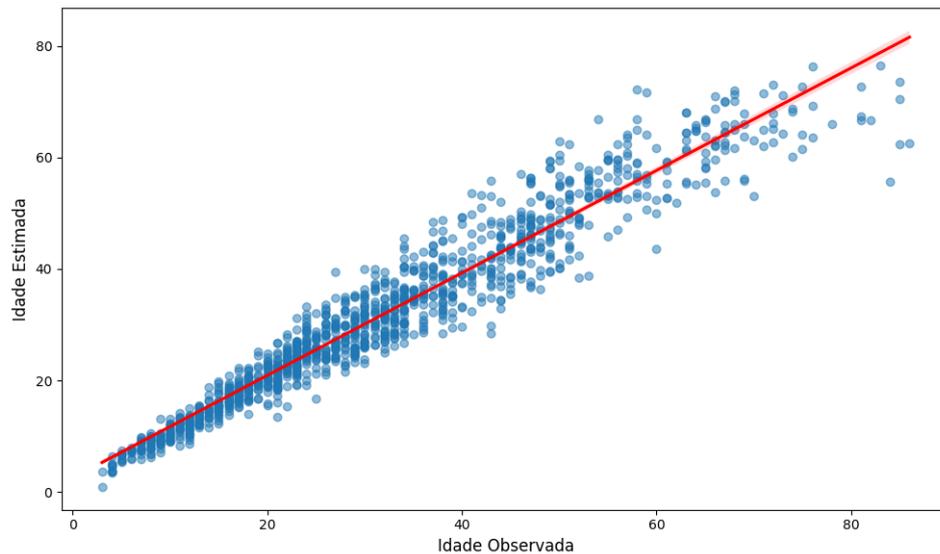


**Figura 4.4** A curva de perda no ajuste fino do framework proposto nos cinco *fold*s utilizados na validação cruzada.



**Figura 4.5** Evolução dos coeficientes de ponderação que se ajustam dinamicamente à importância relativa das perdas durante o ajuste fino (a) coeficiente da tarefa da classificação do sexo, (b) coeficiente da estimação da idade.

Em ambos os gráficos, as curvas mantêm padrões consistentes entre os *fold*s, o que reforça a robustez do processo de validação cruzada e a ausência de variações significativas entre os subconjuntos. No entanto, a Figura 4.5.b, que se refere à tarefa de regressão (idade), apresenta maior oscilação ao longo do treinamento. Isso sugere que o modelo enfrentou maior dificuldade para estabilizar o equilíbrio entre as diferentes perdas envolvidas na tarefa de predição contínua, exigindo maior refinamento na convergência.



**Figura 4.6** Estimativas de idade média dos modelos e seus valores observados. Cada ponto é uma amostra do conjunto de dados, o eixo x representa a idade real de uma amostra, e o eixo y representa a idade que o modelo estimou para essa amostra. A linha diagonal é a linha de estimativas perfeitas. Como pode ser visto, amostras mais jovens estão mais próximas da linha diagonal, enquanto amostras mais velhas têm um grau maior de variância. Também pode ser visto que o modelo não superestima ou subestima sistematicamente a idade.

## 4.5 RESULTADOS EXPERIMENTAIS

### 4.5.1 Estimativa da idade

Na Figura 4.6 ilustra a relação entre a idade observada e a idade estimada pelo modelo proposto, com uma linha vermelha representando a reta ideal em que ambas as idades seriam iguais. Observa-se uma alta correlação entre as idades estimadas e as idades observadas, com um  $R^2$  de 0,9199, indicando que 91,99% da variância nos dados de idade observada é explicada pelo modelo. A concentração da maioria dos pontos ao redor da linha vermelha evidencia a precisão do modelo em realizar previsões próximas aos valores observados. No entanto, nota-se uma leve dispersão para idades mais avançadas, sugerindo maior variabilidade nas estimativas para essa faixa etária.

A Tabela 4.4 apresenta uma síntese dos resultados do erro absoluto médio, mediano e o desvio padrão por faixa etária no conjunto de teste, destacando o desempenho do modelo em diferentes grupos de idade. O erro absoluto médio geral foi de 3,393 e um desvio padrão 3,218, indicando que, em média, as previsões do modelo diferem em cerca de 3,4 anos em relação à idade real, com uma variação considerável entre as amostras. Nas faixas etárias mais jovens, como na faixa etária de 1 a 10 e 11 a 20 anos, o erro absoluto médio foi significativamente menor, com valores de 1,318 e 1,762, respectivamente. Enquanto pior desempenho é para as faixas etárias mais avançadas, o desempenho do modelo se deteriora, atingindo um erro absoluto médio de 15,86 e o erro mediano é de para indivíduos entre 81 e 90 anos. Essa elevação expressiva, acompanhada por um desvio padrão de 1,223,

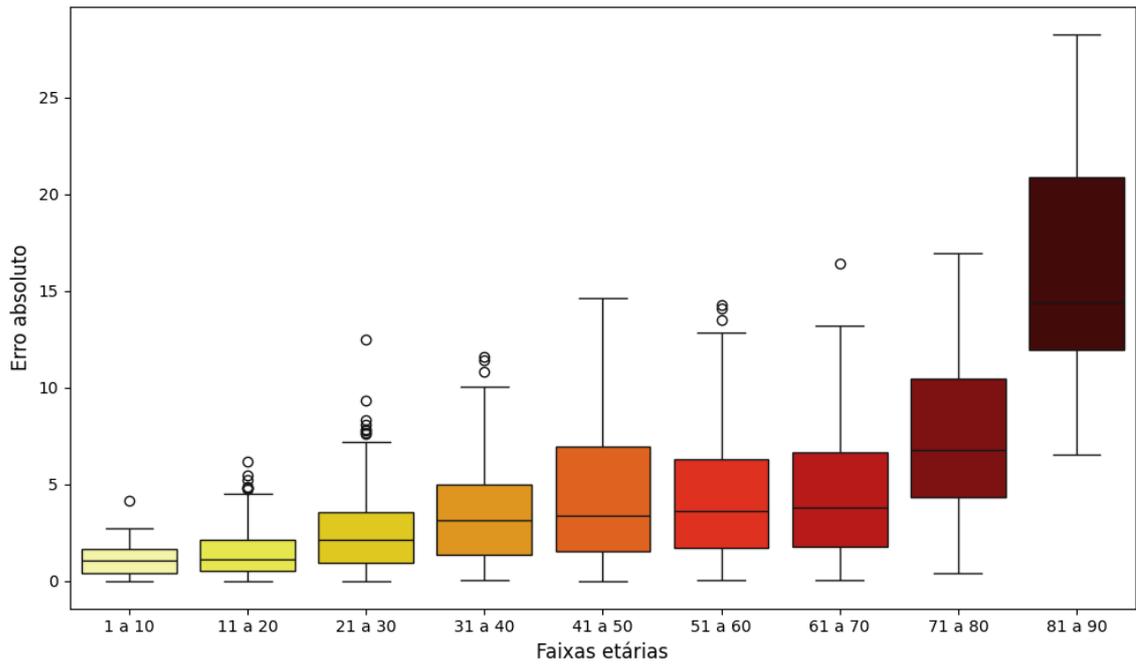
**Tabela 4.4** Os resultados da estimativa de idade por faixa etária no banco de dados de teste. O erro absoluto médio é mostrado como  $\mu$  e a mediana do erro absoluto como  $y$ .

| faixa etária | amostra de teste | $\mu$ | $y$   | desvio padrão |
|--------------|------------------|-------|-------|---------------|
| 1 a 10       | 77               | 1.318 | 1.2   | 0.353         |
| 11 a 20      | 224              | 1.762 | 1.7   | 0.244         |
| 21 a 30      | 319              | 2.74  | 2.66  | 0.316         |
| 31 a 40      | 247              | 3.674 | 3.66  | 0.025         |
| 41 a 50      | 176              | 4.638 | 4.62  | 0.163         |
| 51 a 60      | 109              | 4.664 | 4.66  | 0.091         |
| 61 a 70      | 61               | 4.942 | 5.09  | 0.563         |
| 71 a 80      | 22               | 7.408 | 7.31  | 0.795         |
| 81 a 90      | 10               | 15.86 | 15.71 | 1.223         |
| Total        | 1245             | 3.393 | 2.449 | 3.218         |

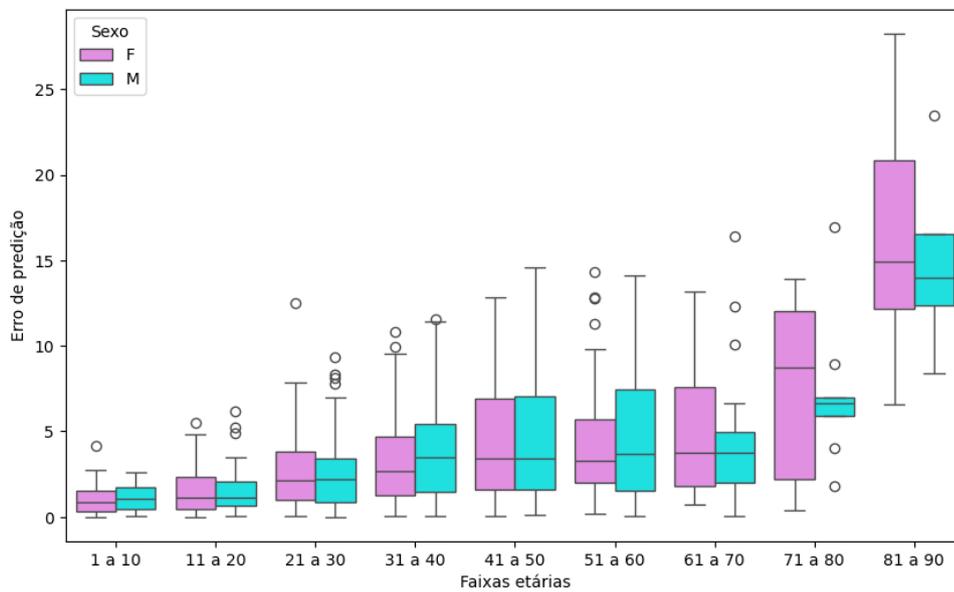
sugere uma alta incerteza das predições para idosos. Observa-se que o erro absoluto médio tende a aumentar conforme a idade dos indivíduos, sugerindo que o modelo encontra maior dificuldade para prever corretamente em idades de pessoas mais velhas. Além do mais, a menor dispersão com o valor de desvio padrão de 0,025 na faixa de 31 a 40 anos sugere que o modelo é mais consistente e apresenta menor variação nas predições para essa faixa etária em comparação com outras, demonstrando maior confiabilidade nas estimativas para esse grupo.

O boxplot apresentado na Figura 4.7 mostra a distribuição do erro absoluto nas predições de idade para diferentes faixas etárias, destacando tendências e variações nos erros do modelo ao longo das idades. Assim como na média, observa-se um aumento gradual nos intervalos interquartis (IQR) à medida que a idade avança, embora a taxa de aumento do IQR não seja uniforme em todas as faixas etárias. Algumas faixas apresentam um aumento mais acentuado, sugerindo a influência de outros fatores além da idade. Nas faixas etárias mais jovens, como de 1 a 10 e 11 a 20 anos, o IQR é relativamente baixo, indicando maior consistência e menor variabilidade nos erros de predição. No entanto, há uma quantidade considerável de *outliers* nas faixas de 11 a 20 e 21 a 30 anos. Por outro lado, a partir da faixa etária de 41 a 50 anos e, especialmente, de 81 a 90 anos, observa-se um aumento acentuado no IQR, revelando maior variabilidade nos erros de predição. A faixa de 81 a 90 anos exibe o maior erro absoluto médio e a maior dispersão, indicando elevada incerteza nas predições do modelo para indivíduos dessa idade.

A Figura 4.8 apresenta outro boxplot que ilustra a distribuição dos erros de predição do modelo para diferentes faixas etárias, segmentada por sexo (feminino e masculino), permitindo uma análise detalhada das diferenças no desempenho do modelo entre os sexos ao longo das faixas etárias. Em algumas faixas etárias, o IQR é maior para homens, enquanto em outras, é maior para mulheres, sugerindo que o impacto da idade no erro de predição varia entre os sexos. Em faixas etárias mais avançadas, especialmente entre 41 a 50 e 71 a 80 anos, o modelo começa a apresentar uma maior variação no erro de predição, com diferenças perceptíveis entre homens e mulheres. Isso indica que o modelo pode ser mais sensível a variações de características em função do sexo em faixas etárias



**Figura 4.7** Gráfico de boxplot dos resultados do erro absoluto na estimativa de idade por faixa etária.



**Figura 4.8** Gráfico de boxplot dos resultados do erro absoluto na estimativa de idade por faixa etária e sexo.

específicas.

**Tabela 4.5** Os resultados da classificação do sexo por faixa etária e pelo sexo no banco de dados de teste.

| Faixa etária | Quantidade | F1-score |           |        |
|--------------|------------|----------|-----------|--------|
|              |            | Feminino | Masculino | Ambos  |
| 1-10         | 77         | 0.762    | 0.947     | 0.751  |
| 11-20        | 224        | 0.978    | 0.951     | 0.937  |
| 21-30        | 319        | 0.994    | 0.971     | 0.969  |
| 31-40        | 247        | 0.999    | 0.963     | 0.969  |
| 41- 50       | 176        | 0.984    | 0.964     | 0.954  |
| 51-60        | 109        | 0.969    | 0.971     | 0.941  |
| 61-70        | 61         | 0.986    | 0.971     | 0.961  |
| 71-80        | 22         | 0.976    | 0.977     | 0.955  |
| 81-91        | 10         | 0.947    | 0.857     | 0.839  |
| Total        | 1245       | 0.977    | 0.962     | 0.9421 |

#### 4.5.2 Classificação do sexo

A Tabela 4.5 mostra os valores médios de f1-score para o modelo proposto em diferentes faixas etárias, discriminando os resultados por sexo (feminino e masculino) e apresentando a combinação em ambos os sexos. O modelo alcançou um f1-score geral de 0,9308 (93,08%), evidenciando um desempenho alto e consistente em quase todas as faixas etárias, com valores médios acima de 0,940 para ambos os sexos e a média combinada. O melhor desempenho foi observado nas faixas etárias intermediárias, especialmente entre 31 e 40 anos, onde o modelo obteve f1-scores de 0,999 para o sexo feminino. Enquanto o melhor desempenho para homens foi na faixa de 71 a 80 anos, com um f1-score de 0,977. Em contrapartida, o desempenho foi ligeiramente inferior nas faixas etárias extremas, particularmente de 1 a 10 anos e de 81 a 91 anos, onde o f1-score combinado foi de 0,751 e 0,839, respectivamente — os menores valores entre todas as faixas.

#### 4.5.3 Estudo ablativo

Os resultados apresentados nas Tabela 4.6 foram obtidas por meio do mesmo protocolo de avaliação e banco de dados, descritos na Seção 4.1.2.3 e na Seção 4.1.1.2, e visam avaliar o impacto de diferentes configurações de arquitetura e ajustes sobre o desempenho do modelo. Primeiramente, a arquitetura EfficientNetV2-L foi utilizada como extração de características das imagens, visando a estimativa de sexo e idade. Diferente do MAE, que emprega a arquitetura ViT, a EfficientNetV2-L utiliza redes convolucionais. Embora a EfficientNetV2-L pré-treinada no conjunto de dados ImageNet tenha sido avaliada com sucesso em tarefas de classificação de sexo (HOUGAZ et al., 2023), estimação da idade (LIANG; CURY; OLIVEIRA, 2023) e em multitarefa de estimação de sexo e idade (PRADO et al., 2024), neste estudo apresentou um desempenho inferior na estimativa de idade, com erro absoluto médio de 4,206, e desempenho bem baixo na classificação de sexo, com um f1-score de 0,5574. A discrepância no desempenho do mo-

**Tabela 4.6** Comparação das arquiteturas ViT-L/16 + KAN, ViT-L/16 e EfficientNetV2 na classificação de sexo e estimativa de idade em modelos multitarefas, a média é mostrado como  $\mu$ .

| Arquitetura           | Estim. idade<br>( $\mu$ erro absoluto médio) | Class. sexo<br>( $\mu$ f1-score) |
|-----------------------|--|----------------------------------|
| EfficientNetV2-L      | 4.206  | 0.5574                           |
| ViT-L/16 (MAE) + MLP  | 3.789  | 0.9338                           |
| <b>MT-DWL ViT-KAN</b> | <b>3.393</b>                                 | <b>0.9421</b>                    |

delo EfficientNetV2-L entre as diferentes tarefas sugere que a arquitetura, apesar de sua eficiência, não generalizou de forma uniforme para todos os problemas. Muitos fatores podem ter influenciado significativamente esses resultados, apenas análises mais aprofundadas posteriores, como a visualização das ativações e a análise de erros, podem fornecer *insights* sobre as limitações do modelo nesse experimento.

Ao incorporar a estratégia de MAE, houve uma melhora significativa no desempenho do modelo. O erro absoluto médio foi reduzido em 0,417 anos, alcançando 3,789, enquanto o f1-score aumentou substancialmente para 0,9338, indicando que a reconstrução de características por meio do ViT permitiu ao modelo capturar melhor os padrões dos dados, resultando em previsões mais precisas. Por conseguinte, com a adição da KAN na última camada, obteve-se o melhor desempenho entre as configurações testadas, com um erro absoluto médio de 3,393 anos, uma redução de 0,396 anos em relação ao ViT isolado, e um aumento marginal no f1-score para 0,9421, mostrando um incremento de 0,0083. Esses resultados sugerem que o MT-DWL ViT-KAN aprimorou a capacidade do modelo de generalizar para os dados de teste, considerando o conjunto de dados em questão.



## **DISCUSSÃO E CONCLUSÃO**

Neste capítulo, são discutidos os resultados obtidos nos experimentos de estimação de idade e classificação de sexo, além de destacar tanto os pontos fortes e as limitações dos métodos propostos. A análise busca contextualizar o desempenho do modelo em relação ao estado da arte, bem como explorar os desafios enfrentados e as oportunidades para aprimoramentos futuros.

### **5.1 ESTIMATIVA DA IDADE**

Conforme indicado na Tabela 4.2, a composição dos dados é predominantemente formada por amostras de indivíduos em faixas etárias jovens e adultas, com uma representação consideravelmente menor de indivíduos nas faixas acima de 51 anos. Essa distribuição desigual impacta diretamente a avaliação do modelo, uma vez que a menor quantidade de amostras em faixas etárias mais avançadas pode prejudicar a capacidade do modelo de generalizar corretamente para esses grupos. A diferença no número de amostras entre faixas etárias induz um viés no processo de otimização, que tende a favorecer as faixas etárias mais representadas, levando o modelo a subestimar a idade em casos de incerteza para minimizar o erro geral. Esse viés é refletido nos resultados da Tabela 4.4, que mostram um aumento progressivo no erro absoluto médio com o avanço da idade, indicando a maior dificuldade do modelo em prever a idade de indivíduos mais velhos.

Embora o aumento no erro absoluto médio com o envelhecimento seja evidente, ele não ocorre de forma linear, apresentando saltos significativos em determinadas faixas. Três aumentos expressivos foram identificados ao transitar entre faixas: de 31 a 40 anos, houve um incremento de 0,93 anos; de 71 a 80 anos, o erro médio aumentou em 2,466 anos; e na faixa de 81 a 90 anos, o erro mais que dobrou, com um aumento de 8,452 anos em relação à faixa anterior. Esse aumento no erro pode ser atribuído a uma combinação de fatores, incluindo o viés dos dados e a ausência de marcadores dentários específicos em idades avançadas. Após os 25 anos, quando a formação dos dentes permanentes se completa, ocorrem poucas alterações anatômicas associadas à maturação dentária. Isso reduz a disponibilidade de indicadores para estimar a idade em radiografias, prejudicando

a precisão do modelo para faixas mais avançadas (PRIETO et al., 2005). Essa limitação indica que a correlação entre idade dentária e idade cronológica diminui substancialmente com o tempo, dificultando a diferenciação entre indivíduos mais velhos. Esse desafio também foi observado em trabalhos anteriores (VILA-BLANCO et al., 2020; MILOŠEVIĆ et al., 2022; LIANG; CURY; OLIVEIRA, 2023), reforçando a necessidade de um conjunto de dados mais equilibrado e de técnicas que incorporem outros indicadores fisiológicos relevantes para essas faixas etárias.

Para minimizar o viés decorrente da distribuição desigual dos dados entre faixas etárias, algumas abordagens poderiam ser implementadas. Uma solução imediata é a ampliação do conjunto de dados, especialmente para as faixas etárias menos representadas, o que permitiria ao modelo uma exposição equilibrada a diferentes idades e reduziria a tendência de subestimar a idade em grupos mais velhos. Outra possibilidade seria a aplicação de uma amostragem estratificada que consiste em dividir o conjunto de dados em subgrupos com base nas faixas etárias e, em seguida, selecionar amostras de cada estrato de forma proporcional para garantir uma representação equilibrada durante o treinamento do modelo. Dessa forma, cada faixa etária seria adequadamente representada no processo de aprendizado, reduzindo a tendência do modelo de favorecer grupos majoritários.

As Figuras 4.7 e 4.8 evidenciam que, na faixa etária de 81 a 90 anos, o erro de predição para mulheres é significativamente maior do que para homens, com maior dispersão. Esse comportamento pode indicar dificuldades do modelo em generalizar para mulheres mais velhas, devido à menor representatividade de dados ou a variações fisiológicas mais marcantes no envelhecimento feminino. E também, a alta concentração de *outliers* em algumas faixas etárias sugere a influência de fatores adicionais, como características anatômicas ou fisiológicas específicas a subgrupos populacionais. Uma análise detalhada desses *outliers* pode revelar causas subjacentes e contribuir para melhorias no desempenho do modelo.

## 5.2 CLASSIFICAÇÃO DO SEXO

A precisão na classificação do sexo seguiu um padrão distinto em relação ao desempenho da estimativa de idade. O menor f1-score foi registrado tanto em crianças (1 a 10 anos) quanto em idosos (acima de 80 anos), destacando os desafios na predição para esses grupos extremos, enquanto a precisão foi maior em pacientes adultos, corroborando achados de estudos anteriores (ILIĆ; VODANOVIĆ; SUBAŠIĆ, 2019; PRADO et al., 2024). Em idades muito jovens, as diferenças morfológicas são menos pronunciadas devido à influência hormonal, com mudanças mais acentuadas após a puberdade, especialmente em mulheres que completam esse desenvolvimento mais cedo (KRENZER, 2006). Trabalhos como Cuenca (1994) e (SCHUTKOWSKI, 1993) apontam critérios morfológicos importantes em crianças, como o pronunciamento do queixo, mais largo e quadrangular nos meninos, e a eversão da região gônica, que é plana nas meninas e mais saliente nos meninos. Enquanto em idades mais avançadas, mudanças fisiológicas associadas ao envelhecimento, como a masculinização das características cranianas em mulheres após a menopausa, e a perda de robustez óssea nos homens (KRENZER, 2006). Do mesmo modo, a menor quantidade

de amostras nessas faixas etárias reduz a representatividade dos dados, dificultando a eficácia do modelo.

O f1-score para o sexo feminino (0,977) foi ligeiramente superior ao do masculino (0,962), indicando que o modelo pode estar um pouco mais ajustado para identificar características associadas ao sexo feminino, conforme também observado no estudo de Ciconelle et al. (2023). Essa diferença pode ser parcialmente explicada pela distribuição das amostras, já que no trabalho de Ciconelle et al. (2023), 58% das amostras pertenciam ao sexo feminino, enquanto no presente estudo, essa proporção é de 60,13%. Apesar de pequena, essa discrepância sugere que o balanceamento das amostras pode ser um fator para garantir um desempenho consistente entre os sexos e em todas as faixas etárias.

### 5.3 PONTOS FORTES E LIMITAÇÕES

A metodologia proposta, que combina a arquitetura ViT pré-treinada com MAE e o ajuste fino utilizando KAN em um cenário multitarefa, demonstrou ser uma abordagem promissora. Os experimentos evidenciaram melhorias no desempenho, com redução do erro absoluto médio e aumento do f1-score em comparação ao modelo EfficientNet-Large. Os resultados reforçam o potencial de técnicas emergentes de aprendizado não supervisionado para extrair características relevantes de imagens médicas, mesmo diante de conjuntos de dados limitados ou desbalanceados. Além disso, a abordagem multitarefa permite a estimação simultânea de sexo e idade, maximizando a eficiência dos recursos computacionais. Em um contexto onde as tarefas se complementam, essa abordagem permite a identificação de padrões e relações ocultas, resultantes do aprendizado compartilhado entre as diferentes tarefas, que poderiam ser negligenciados em um ambiente de aprendizado focado em apenas uma tarefa.

Ademais, é fundamental considerar certas limitações. Fatores como o tamanho das amostras, a variabilidade biológica inerente e a diversidade populacional desempenham um papel central na análise. Diferenças morfológicas significativas podem ser observadas no esqueleto e nos dentes, variando substancialmente entre grupos geográficos e étnicos (HERRERA-ESCUADERO; TORO; PARADA-SANCHEZ, 2024; KRENZER, 2006). Além disso, influências socioecológicas, como nutrição, clima e patologias, afetam diretamente o desenvolvimento ósseo e, conseqüentemente, a morfologia esquelética (KRENZER, 2006). Esses fatores ressaltam a complexidade das tarefas e indicam a necessidade de abordagens que levem em conta a diversidade das populações.

### 5.4 TRABALHOS FUTUROS

Para o direcionamento de pesquisas futuras, algumas questões podem ser investigadas. Entre elas, destaca-se o impacto do treinamento com um volume maior de dados e a avaliação da utilização conjunta de radiografias panorâmicas e cefalométricas, ou até mesmo de radiografias de outras regiões do esqueleto. Essa abordagem é promissora, especialmente considerando o estudo de Capitaneanu, Willems e Thevissen (2017), que menciona uma pesquisa que, ao integrar parâmetros cefalométricos e odontométricos, elevou a precisão na determinação do sexo em comparação com a análise isolada desses

fatores. E ainda, recomenda-se uma investigação aprofundada sobre a otimização dos hiperparâmetros de treinamento e a experimentação com novas funções de custo, como a normalização de gradiente para perdas adaptativas. Outra perspectiva interessante seria o desenvolvimento de uma arquitetura híbrida que integre CNNs e ViTs com a técnica MAE, permitindo a extração simultânea de características globais e locais das imagens, conforme discutido por Yeom et al. (2023).

## REFERÊNCIAS BIBLIOGRÁFICAS

ADSERIAS-GARRIGA, J. et al. When forensic odontology met biochemistry: Multidisciplinary approach in forensic human identification. *Archives of Oral Biology*, Pergamon, v. 87, p. 7–14, 3 2018. ISSN 0003-9969.

ALQAHTANI, S. J.; HECTOR, M. P.; LIVERSIDGE, H. M. Brief communication: The London atlas of human tooth development and eruption. *American journal of physical anthropology*, Am J Phys Anthropol, v. 142, n. 3, p. 481–490, jul 2010. ISSN 1096-8644. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/20310064/>>.

ANDERSON, D.; THOMPSON, G. Interrelationships and sex differences of dental and skeletal measurements. *Journal of Dental Research*, v. 52, p. 8–431, 1973.

ANDRADE, K. M. et al. Automatic dental biofilm detection based on deep learning. *Journal of Clinical Periodontology*, John Wiley and Sons, Ltd, v. 50, p. 571–581, 5 2023. ISSN 1600-051X. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1111/jcpe.13774><https://onlinelibrary.wiley.com/doi/abs/10.1111/jcpe.13774><https://onlinelibrary.wiley.com/doi/10.1111/jcpe.13774>>.

BADRAN, D. H. et al. Predictive accuracy of mandibular ramus flexure as a morphologic indicator of sex dimorphism in jordanians. *International Journal of Morphology*, v. 33, n. 4, p. 1248–1254, 2015.

BERMAN, G. et al. Dental identification. *Manual of Forensic Odontology, Fifth Edition*, CRC Press, p. 75–128, 1 2013. Disponível em: <<http://www.crcnetbase.com/doi/10.1201/b13744-5>>.

BROWN, T. B. et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, Neural information processing systems foundation, v. 2020-December, 5 2020. ISSN 10495258. Disponível em: <<https://arxiv.org/abs/2005.14165v4>>.

CAMERIERE, R.; FERRANTE, L.; CINGOLANI, M. Age estimation in children by measurement of open apices in teeth. *International journal of legal medicine*, Int J Legal Med, v. 120, n. 1, p. 49–53, jan 2006. ISSN 0937-9827. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/16283352/>>.

CAPITANEANU, C.; WILLEMS, G.; THEVISSSEN, P. A systematic review of odontological sex estimation methods. *The Journal of Forensic Odonto-Stomatology - JFOS*, v. 35, p. 1: 19–1: 19, 12 2017. ISSN 2219-6749. Disponível em: <<https://ojs.iofos.eu/index.php/Journal/article/view/1601>>.

- CHU, P. et al. Using Octuplet Siamese Network for Osteoporosis Analysis on Dental Panoramic Radiographs. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Institute of Electrical and Electronics Engineers Inc., v. 2018-July, p. 2579–2582, oct 2018. ISSN 1557170X.
- CICONELLE, A. C. M. et al. Deep learning for sex determination: Analyzing over 200,000 panoramic radiographs. *Journal of Forensic Sciences*, v. 68, n. 6, p. 2057–2064, 2023. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/1556-4029.15376>>.
- CIPOLLA, R.; GAL, Y.; KENDALL, A. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, p. 7482–7491, 5 2017. ISSN 10636919. Disponível em: <<https://arxiv.org/abs/1705.07115v3>>.
- CLAUDIO, I. et al. Evaluating accuracy and precision in morphologic traits for sexual dimorphism in malnutrition human skull: a comparative study evaluación de la exactitud y precisión en los rasgos morfológicos de dimorfismo sexual en cráneos humanos malnutridos: un estudio comparativo. *Int. J. Morphol*, v. 26, p. 877–881, 2008.
- CRAWSHAW, M. Multi-task learning with deep neural networks: A survey. *CoRR*, abs/2009.09796, 2020. Disponível em: <<https://arxiv.org/abs/2009.09796>>.
- CUENCA, J. *Introducción a la antropología forense: análisis e identificación de restos óseos humanos*. Anaconda, 1994. ISBN 9789589200018. Disponível em: <<https://books.google.com.co/books?id=RU4nMwEACAAJ>>.
- CULAR, L. et al. Dental age estimation from panoramic X-ray images using statistical models. *International Symposium on Image and Signal Processing and Analysis, ISPA*, IEEE Computer Society, p. 25–30, oct 2017. ISSN 18492266.
- DEMIRJIAN, A.; GOLDSTEIN, H. New systems for dental maturity based on seven and four teeth. *Annals of human biology*, Ann Hum Biol, v. 3, n. 5, p. 411–421, 1976. ISSN 0301-4460. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/984727/>>.
- DOSOVITSKIY, A. et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. 2021.
- ECKERT, W. G.; GARLAND, N. The history of the forensic applications in radiology. *The American journal of forensic medicine and pathology*, Am J Forensic Med Pathol, v. 5, p. 53–56, 1984. ISSN 0195-7910. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/6369961/>>.
- GUSTAFSON, G. Age determinations on teeth. *The Journal of the American Dental Association*, Elsevier, v. 41, p. 45–54, 7 1950. ISSN 0002-8177.
- HE, Y. et al. Age and sex estimation in cephalometric radiographs based on multitask convolutional neural networks. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, Mosby, v. 138, p. 225–231, 7 2024. ISSN 2212-4403.

HEINRICH, A. Accelerating computer vision-based human identification through the integration of deep learning-based age estimation from 2 to 89 years. *Scientific Reports* 2024 14:1, Nature Publishing Group, v. 14, p. 1–11, 2 2024. ISSN 2045-2322. Disponível em: <<https://www.nature.com/articles/s41598-024-54877-1>>.

HENG, D.; MANICA, S.; FRANCO, A. Forensic dentistry as an analysis tool for sex estimation: A review of current techniques. *Research and Reports in Forensic Medical Science*, Dove Press, v. 12, p. 25–39, 12 2022. Disponível em: <<https://www.dovepress.com/forensic-dentistry-as-an-analysis-tool-for-sex-estimation-a-review-of-peer-reviewed-fulltext-article-1>>

HERRERA-ESCUADERO, T. M.; TORO, D. A.; PARADA-SANCHEZ, M. T. How teeth can be used to estimate sexual dimorphism? a scoping review. *Forensic Science International*, Elsevier, v. 360, p. 112061, 7 2024. ISSN 0379-0738.

HIRUNCHAVAROD, N. et al. Deeptoothduo: Multi-task age-sex estimation and understanding via panoramic radiograph. 2024 *IEEE International Symposium on Biomedical Imaging (ISBI)*, IEEE, p. 1–5, 5 2024. Disponível em: <<https://ieeexplore.ieee.org/document/10635634/>>.

HOUGAZ, A. et al. Sex estimation on panoramic dental radiographs: A methodological approach. In: *Anais do XXIII Simpósio Brasileiro de Computação Aplicada à Saúde*. Porto Alegre, RS, Brasil: SBC, 2023. p. 115–125. ISSN 2763-8952. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/25282>>.

ILIĆ, I.; VODANOVIĆ, M.; SUBAŠIĆ, M. Gender estimation from panoramic dental x-ray images using deep convolutional networks. *IEEE*, p. 1–5, 2019.

JADER, G. et al. Deep instance segmentation of teeth in panoramic x-ray images. In: *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.: s.n.], 2018. p. 400–407.

KARAMAN, F. Use of diagonal teeth measurements in predicting gender in a turkish population\*. *Journal of Forensic Sciences*, John Wiley & Sons, Ltd, v. 51, p. 630–635, 5 2006. ISSN 1556-4029. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1111/j.1556-4029.2006.00133.x><https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1556-4029.2006.00133.x><https://onlinelibrary.wiley.com/doi/10.1111/j.1556-4029.2006.00133.x>>.

KERMANI, M.; YAZDI, F. T.; HAGHIGHI, M. A. Evaluation of the accuracy of demirjian's method for estimating chronological age from dental age in shiraz, iran: Using geometric morphometrics method. *Clinical and Experimental Dental Research*, v. 5, n. 3, p. 191–198, 2019. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/cre2.169>>.

KRENZER, U. I. Compendio de métodos antropológico forenses para la reconstrucción del perfil osteo-biológico cafa. 2006.

KUCHLER, E. C. et al. Mandibular and dental measurements for sex determination using machine learning. *Scientific Reports* 2024 14:1, Nature Publishing Group, v. 14, p. 1–9, 4 2024. ISSN 2045-2322. Disponível em: <<https://www.nature.com/articles/s41598-024-59556-9>>.

KUO, Y. F. et al. A convolutional neural network approach for dental panoramic radiographs classification. *Journal of Medical Imaging and Health Informatics*, American Scientific Publishers, v. 7, n. 8, p. 1693–1704, dec 2017. ISSN 21567026.

KVAAL, S. I. et al. Age estimation of adults from dental radiographs. *Forensic Science International*, Forensic Sci Int, v. 74, n. 3, p. 175–185, jul 1995. ISSN 03790738. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/7557754/>>.

LAMENDIN, H. et al. A simple technique for age estimation in adult corpses: The two criteria dental method. *Journal of Forensic Sciences*, American Society of Mechanical Engineers Digital Collection, v. 37, p. 1373–1379, 9 1992. ISSN 0022-1198. Disponível em: <<https://dx.doi.org/10.1520/JFS13327J>>.

LEWIS, J. M. et al. - dental age estimation. *Manual of Forensic Odontology, Fifth Edition*, CRC Press, p. 228–273, 1 2013. Disponível em: <<https://www.taylorfrancis.com/chapters/edit/10.1201/b13744-13/dental-age-estimation-james-lewis-david-senn>>.

LIANG, J.; CURY, P.; OLIVEIRA, L. Revisiting age estimation on panoramic dental images. In: *2023 36th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.: s.n.], 2023. p. 205–210.

LIU, Z. et al. Kan: Kolmogorov-arnold networks. 4 2024. Disponível em: <<https://arxiv.org/abs/2404.19756v4>>.

LOTH, S. R.; HENNEBERG, M. Mandibular ramus flexure: A new morphologic indicator of sexual dimorphism in the human skeleton. *American Journal of Physical Anthropology*, v. 99, n. 3, p. 473–485, 1996. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/\%28SICI\%291096-8644\%28199603\%2999\%3A3\%3C473\%3A\%3AAID-AJPA8\%3E3.0.CO\%3B2-X>>.

MAPLES, W. An improved technique using dental histology for estimation of adult age. *Journal of Forensic Sciences*, American Society of Mechanical Engineers Digital Collection, v. 23, p. 764–770, 10 1978. ISSN 0022-1198. Disponível em: <<https://dx.doi.org/10.1520/JFS10735J>>.

MILOŠEVIĆ, D. et al. Automated estimation of chronological age from panoramic dental X-ray images using deep learning. *Expert Systems with Applications*, Pergamon, v. 189, p. 116038, mar 2022. ISSN 0957-4174.

MILOŠEVIĆ, D. et al. Estimating biological gender from panoramic dental x-ray images. *International Symposium on Image and Signal Processing and Analysis, ISPA*, v. 2019-September, p. 105–110, 2019.

MILOŠEVIĆ, D. et al. A comprehensive exploration of neural networks for forensic analysis of adult single tooth x-ray images. *IEEE*, p. 70980–71002, 2022.

MOHAMED, E. G. et al. Dental age estimation using deep learning: A comparative survey. *Computation*, v. 11, n. 2, 2023. ISSN 2079-3197. Disponível em: <<https://www.mdpi.com/2079-3197/11/2/18>>.

MOORREES, C. F.; FANNING, E. A.; HUNT, E. E. Formation and resorption of three deciduous teeth in children. *American Journal of Physical Anthropology*, John Wiley & Sons, Ltd, v. 21, p. 205–213, 6 1963. ISSN 1096-8644. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1002/ajpa.1330210212https://onlinelibrary.wiley.com/doi/abs/10.1002/ajpa.1330210212https://onlinelibrary.wiley.com/doi/10.1002/ajpa.1330210212>>.

NAGARE, S. P.; CHAUDHARI R. S.AND BIRANGANE, R. S.; PARKARWAR, P. C. Sex determination in forensic identification, a review. *Journal of forensic dental sciences*, [https://doi.org/10.4103/jfo.jfds\\_55\\_17](https://doi.org/10.4103/jfo.jfds_55_17), v. 10, n. 2, p. 61–66, 2018.

OLIVEIRA, W. et al. Estimation of human age using machine learning on panoramic radiographs for brazilian patients. *Scientific Reports 2024 14:1*, Nature Publishing Group, v. 14, p. 1–16, 8 2024. ISSN 2045-2322. Disponível em: <<https://www.nature.com/articles/s41598-024-70621-1>>.

PINHEIRO, L. et al. Numbering permanent and deciduous teeth via deep instance segmentation in panoramic x-rays. In: RITTNER, L. et al. (Ed.). *17th International Symposium on Medical Information Processing and Analysis*. SPIE, 2021. v. 12088, n. 120880C, p. 1–10. Disponível em: <<https://doi.org/10.1117/12.2606211>>.

POEDJIASTOETI, W.; SUEBNUKARN, S. Application of Convolutional Neural Network in the Diagnosis of Jaw Tumors. *Healthcare informatics research*, Healthc Inform Res, v. 24, n. 3, p. 236–241, jul 2018. ISSN 2093-3681. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/30109156/>>.

PRADO, I. et al. Multi-task learning based on log dynamic loss weighting for sex classification and age estimation on panoramic radiographs. In: *2024 19th International Conference on Computer Vision Theory and Applications (VISAPP)*. Roma, Italy: [s.n.], 2024.

PRIETO, J. L. et al. Evaluation of chronological age based on third molar development in the spanish population. *International Journal of Legal Medicine*, Springer, v. 119, p. 349–354, 11 2005. ISSN 09379827. Disponível em: <<https://link.springer.com/article/10.1007/s00414-005-0530-3>>.

RAJEE, M.; MYTHILI, C. Gender classification on digital dental x-ray images using deep convolutional neural network. *Biomedical Signal Processing and Control*, v. 69, p. 102–139, 2021.

RAJENDRAN, A.; SIVAPATHASUNDHARAM, B. *Shafer's Textbook of Oral Pathology*. Elsevier Health Sciences, 2012. ISBN 9788131238004. Disponível em: <<https://books.google.com.br/books?id=WnhtAwAAQBAJ>>.

SAINI, V. et al. Mandibular ramus: An indicator for sex in fragmentary mandible\*. *Journal of Forensic Sciences*, John Wiley Sons, Ltd, v. 56, p. S13–S16, 1 2011. ISSN 1556-4029. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1111/j.1556-4029.2010.01599.x><https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1556-4029.2010.01599.x><https://onlinelibrary.wiley.com/doi/10.1111/j.1556-4029.2010.01599.x>>.

SAUNDERS, E. The teeth a test of age, considered with reference to the factory children. *The American Journal of Dental Science*, Biomedical Journal Digitization Project, v. 7, p. 330, 6 1847. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6080001/>>.

SCHMELING, A. et al. Forensic age estimation: Methods, certainty, and the law. *Deutsches Arzteblatt International*, Deutscher Arzte-Verlag GmbH, v. 113, n. 4, p. 44, jan 2016. ISSN 18660452.

SCHOUR, I.; MASSLER, M. The development of human dentition. *Chicago: American Dental Association*, 1941.

SCHUTKOWSKI, H. Sex determination of infant and juvenile skeletons: I. morphognostic features. *American Journal of Physical Anthropology*, John Wiley and Sons, Ltd, v. 90, p. 199–205, 2 1993. ISSN 1096-8644. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1002/ajpa.1330900206><https://onlinelibrary.wiley.com/doi/abs/10.1002/ajpa.1330900206><https://onlinelibrary.wiley.com/doi/10.1002/ajpa.1330900206>>.

SENER, O.; KOLTUN, V. Multi-task learning as multi-objective optimization. *Advances in Neural Information Processing Systems*, Neural information processing systems foundation, v. 2018-December, p. 527–538, 10 2018. ISSN 10495258. Disponível em: <<https://arxiv.org/abs/1810.04650v2>>.

SILVA, G.; OLIVEIRA, L.; PITHON, M. Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives. *Expert Systems with Applications*, Pergamon, v. 107, p. 15–31, oct 2018. ISSN 0957-4174.

SOLHEIM, T. Dental age estimation. an alternative technique for tooth sectioning. *The American journal of forensic medicine and pathology*, v. 5, n. 2, p. 181–184, June 1984. ISSN 0195-7910. Disponível em: <<http://europepmc.org/abstract/MED/6731411>>.

SORENTI, M. et al. Sexual dimorphism of dental tissues in modern human mandibular molars. *American Journal of Physical Anthropology*, v. 169, n. 2, p. 332–340, 2019. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/ajpa.23822>>.

TUZOFF, D. V. et al. Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dento maxillo facial radiology*, Dentomaxillofac Radiol, v. 48, n. 4, 2019. ISSN 0250-832X. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/30835551/>>.

VILA-BLANCO, N. et al. Deep Neural Networks for Chronological Age Estimation from OPG Images. *IEEE Transactions on Medical Imaging*, Institute of Electrical and Electronics Engineers Inc., v. 39, n. 7, p. 2374–2384, jul 2020. ISSN 1558254X.

VILA-BLANCO, N. et al. A systematic overview of dental methods for age assessment in living individuals: from traditional to artificial intelligence-based approaches. *International journal of legal medicine*, Int J Legal Med, v. 137, p. 1117–1146, 7 2023. ISSN 1437-1596. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/37055627/>>.

VILA-BLANCO, N. et al. A systematic overview of dental methods for age assessment in living individuals: from traditional to artificial intelligence-based approaches. *International journal of legal medicine*, Int J Legal Med, v. 137, p. 1117–1146, 7 2023. ISSN 1437-1596. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/37055627/>>.

WOLF, D. et al. Self-supervised pre-training with contrastive and masked autoencoder methods for dealing with small datasets in deep learning for medical imaging. *Scientific Reports 2023 13:1*, Nature Publishing Group, v. 13, p. 1–13, 11 2023. ISSN 2045-2322. Disponível em: <<https://www.nature.com/articles/s41598-023-46433-0>>.

YEOM, H. G. et al. Estimating chronological age through learning local and global features of panoramic radiographs in the korean population. *Scientific Reports 2023 13:1*, Nature Publishing Group, v. 13, p. 1–8, 12 2023. ISSN 2045-2322. Disponível em: <<https://www.nature.com/articles/s41598-023-48960-2>>.