Catalogação na publicação
Seção de Catalogação e Classificação

L732a Lima, Ismael Alves.
Avaliação de arquiteturas de redes neurais convolucionais para predição do prognóstico de crescimento maxilofacial em pacientes com fissuras labiopalatinas / Ismael Alves Lima. - João Pessoa, 2025.
28 f. : il.

Orientação: Thaís Gaudencio do Rêgo, Rosa Helena Wanderley Lacerda.
TCC (Graduação) - UFPB/CI.

1. Fissura labiopalatina. 2. Inteligência artificial. 3. Índice de GOSLON. 4. Rede neural convolucional. 5. Escaneamento intraoral. I. Rêgo, Thaís Gaudencio do. II. Lacerda, Rosa Helena Wanderley. III. Título.

UFPB/CI CDU 004.8

Elaborado por MICHELLE DE KASSIA FONSECA BARBOSA - CRB-738

Avaliação de Arquiteturas de Redes Neurais Convolucionais para Predição do Prognóstico de Crescimento Maxilofacial em Pacientes com Fissuras Labiopalatinas

Ismael Alves Lima[1], Alice Castro Guedes Mendonça[2], Rosa Helena Wanderley Lacerda[2], Thaís Gaudencio do Rêgo[1]

[1]Centro de Informática - Universidade Federal da Paraíba (UFPB) - João Pessoa/PB - Brasil
[2]Centro de Ciências da Saúde - Universidade Federal da Paraíba (UFPB) - João Pessoa/PB - Brasil

ismael.alves@academico.ufpb.br,
{alicecgm1,rhelenawanderley,gaudenciothais}@gmail.com

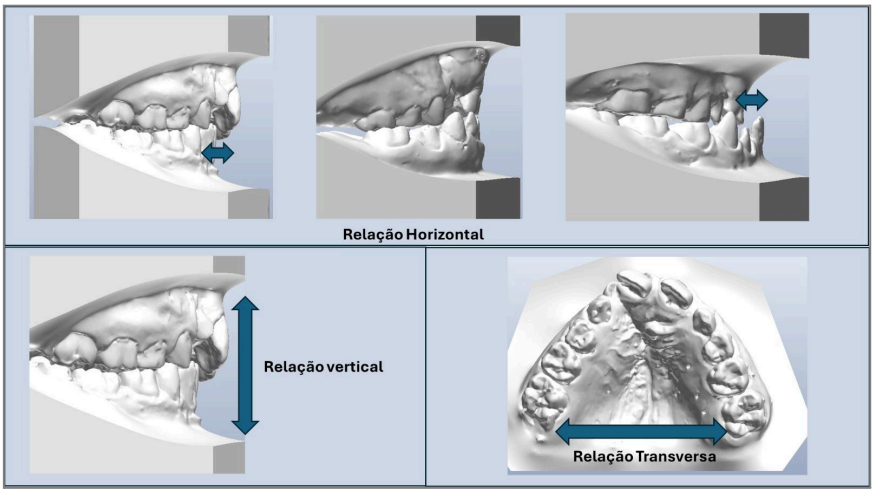
Abstract: *Cleft lip and palate are the most common congenital craniofacial deformities worldwide, affecting approximately 1 in every 1000-1500 births. One of the crucial aspects of the treatment and rehabilitation of cleft patients is the analysis of maxillofacial development, performed through the GOSLON YARDSTICK, which guides surgical planning and allows early diagnosis of possible complications associated with cleft. The main objective of this work is to evaluate the performance of convolutional neural networks in predicting the prognosis of maxillofacial growth in cleft patients, as well as to develop an AI model applicable to clinical practice. This work evaluates the performance of the VGG16, VGG19, ResNet50, InceptionV3, Xception, DenseNet201, EfficientNetV2L and MobileNetV3L networks on a dataset containing 597 intraoral scan images divided into three classes, corresponding to the prognosis of facial growth. The results obtained range from 72% accuracy for EfficientNetV2L to 83.5% for VGG19. To add interpretability to the predictions, the Grad-CAM technique was used, which highlights the most important areas for classification. These results highlight the potential of AI in supporting the treatment and rehabilitation of patients with cleft lip and palate.*

Resumo: *As fissuras labiopalatinas são as deformidades craniofaciais congênitas mais comuns em todo o mundo, afetando cerca de 1 em cada 1000-1500 nascimentos. Um dos aspectos cruciais do tratamento e reabilitação dos pacientes fissurados é a análise do desenvolvimento maxilofacial, realizada através do Índice de GOSLON, que orienta o planejamento cirúrgico e permite o diagnóstico precoce de possíveis complicações associadas à fissura. Este trabalho tem como objetivo principal avaliar o desempenho de redes neurais convolucionais na predição do prognóstico de crescimento maxilofacial em pacientes fissurados, bem como desenvolver um modelo de inteligência artificial aplicável à prática clínica. Este trabalho avalia o desempenho das redes VGG16, VGG19, ResNet50, InceptionV3, Xception, DenseNet201, EfficientNetV2L e MobileNetV3L em um conjunto de dados contendo 597 imagens de escaneamento intraorais divididas em três classes, correspondentes ao prognóstico de crescimento facial. Os resultados obtidos variam entre 72% de acurácia para a EfficientNetV2L e 83,5% para a VGG19. Para adicionar interpretabilidade às previsões, foi utilizada a técnica Grad-CAM, que destaca as áreas mais importantes para a classificação. Estes resultados evidenciam o potencial da IA no apoio ao tratamento e reabilitação de pacientes com fissuras labiopalatinas.*

1. Introdução

As fissuras labiopalatinas constituem as anomalias craniofaciais congênitas com maior prevalência no mundo. A Organização Mundial da Saúde (OMS) (2025) estima uma frequência global média de 1 a cada 1.000-1.500 nascimentos. Entre as causas relacionadas, a predisposição genética é a mais importante, no entanto, a má nutrição durante a gestação, assim como o consumo de tabaco e álcool, também são fatores associados. Para Marazita e Mooney (2004), "as fissuras labiopalatinas representam um importante problema de saúde pública, pois estão associadas a uma maior morbidade e mortalidade, em comparação com indivíduos não afetados". Em razão dessas particularidades, os indivíduos afetados necessitam de acompanhamento multidisciplinar e especializado ao longo de diversas fases da vida, o que envolve várias etapas de tratamento e procedimentos cirúrgicos [3].

Figura 1: Relações horizontais, verticais e transversais interarcos.



Fonte: Autor.

Nesse contexto, a Ortodontia desempenha um papel essencial no tratamento e na reabilitação funcional e estética de pacientes com fissuras labiopalatinas, para os quais foram desenvolvidos procedimentos específicos. Dentre esses, destaca-se a análise oclusal, com o Índice de GOSLON sendo o principal método utilizado para definição da severidade dos casos, prognóstico e estabelecimento de plano de tratamento. Criado por Mars et al. (1987), esse índice tornou-se amplamente difundido por ser um método simples e altamente confiável. Ele classifica a má oclusão em pacientes com fissuras labiopalatinas unilaterais completas, representando tanto a gravidade da alteração

quanto a dificuldade de correção, com base na relação interarcos nas dimensões ântero-posterior, transversal e vertical (Figura 1) [5]..

Em face da complexidade dos casos envolvendo fissuras labiopalatinas, torna-se cada vez mais essencial contar com ferramentas que auxiliem no planejamento e na tomada de decisões clínicas. Nesse contexto, a Inteligência Artificial (IA) vem se destacando como um recurso promissor, tanto na prática ortodôntica e odontológica quanto na clínica em geral. Com os avanços recentes na digitalização de dados, no aprendizado de máquina e na infraestrutura computacional, as aplicações de IA estão se expandindo para áreas que antes eram consideradas exclusivas de especialistas humanos [6].

As aplicações mais recentes da IA em odontologia incluem por exemplo o uso de aprendizado profundo (do inglês, *deep learning*) no diagnóstico de lesões de cárie por imagens radiográficas [7], segmentação dos canais mandibulares através de imagens de tomografia computadorizada [8], e também o diagnóstico de câncer oral através de imagens fotográficas [9]. Nesse contexto, a IA torna-se uma grande aliada da odontologia, sendo utilizada principalmente para aumentar a precisão e a eficiência do diagnóstico, aspectos fundamentais para garantir os melhores resultados nos procedimentos e, ao mesmo tempo, oferecer cuidados de excelência [10].

Considerando a necessidade de uma avaliação precisa, a escassez de profissionais especializados em fissuras labiopalatinas e os avanços recentes na aplicação de inteligência artificial na área da saúde, este estudo tem como objetivo avaliar o desempenho de diferentes arquiteturas de redes neurais convolucionais na predição do prognóstico de crescimento maxilofacial em pacientes com fissuras labiopalatinas, além de desenvolver, com base nos resultados obtidos, um modelo inteligente de apoio ao tratamento desses pacientes.

Para isso, foram utilizadas imagens de escaneamento intraorais e redes neurais convolucionais (do inglês, *Convolutional Neural Network*, CNN) pré-treinadas no conjunto de dados ImageNet[1]. Essas redes foram ajustadas por meio da técnica de

[1] O ImageNet é um banco de dados de imagens organizado de acordo com a hierarquia WordNet (atualmente apenas os substantivos), em que cada nó da hierarquia é representado por centenas e milhares de imagens. O projeto tem sido fundamental no avanço da visão computacional e da pesquisa de aprendizado profundo. Os dados estão disponíveis gratuitamente para pesquisadores para uso não comercial.

fine-tuning, que consiste na adaptação de um modelo previamente treinado a uma nova tarefa, permitindo a reutilização de representações aprendidas e acelerando o processo de aprendizagem em contextos específicos. Adicionalmente, a técnica Grad-CAM foi empregada para destacar as áreas de maior importância na previsão do modelo, proporcionando maior interpretabilidade aos resultados.

2. Trabalhos Relacionados

Após uma revisão da literatura nas bases de dados PubMed e Scielo, não foram encontrados estudos que abordem diretamente a avaliação do prognóstico de crescimento maxilofacial, por meio da inteligência artificial e análise de imagens de escaneamentos intraorais. Portanto, nesta seção serão abordadas as pesquisas que envolveram de alguma forma a aplicação de IA no tratamento de pacientes com fissuras.

É possível destacar, neste âmbito, a revisão da literatura conduzida por Huqh et al (2021), que teve como objetivo identificar estudos sobre a aplicação de técnicas de IA no contexto das fissuras labiopalatinas. A revisão revelou que a maioria das pesquisas concentra-se na detecção da hipernasalidade e na avaliação de sua gravidade [11-15]. Além disso, foram identificados estudos voltados para a avaliação do risco genético [16,17], a automatização da criação de marcadores cirúrgicos [18] e a identificação das características e relações dentárias em crianças com diferentes tipos de fissuras [19,20]. Outros trabalhos exploraram a predição pré-natal das fissuras labiopalatinas, com base em variáveis coletadas das gestantes [21].

Em outro estudo, Huqh et al. (2023) desenvolveram um modelo para a predição e avaliação do crescimento maxilofacial em crianças com fissuras labiopalatinas. O banco de dados foi construído a partir da coleta de 14 características dentárias de 100 pacientes, utilizando radiografias cefalométricas laterais de indivíduos com e sem fissura. Para a análise, foi empregada uma rede neural para classificar o crescimento maxilofacial em três categorias: normal, excessivo e deficiente. Posteriormente, as associações entre as variáveis foram avaliadas por meio de regressão logística ordinal. A abordagem proposta apresentou um Erro Quadrático Médio Previsto (PMSE) de 2,03% e evidenciou uma forte associação entre gênero, idade e a presença da fissura.

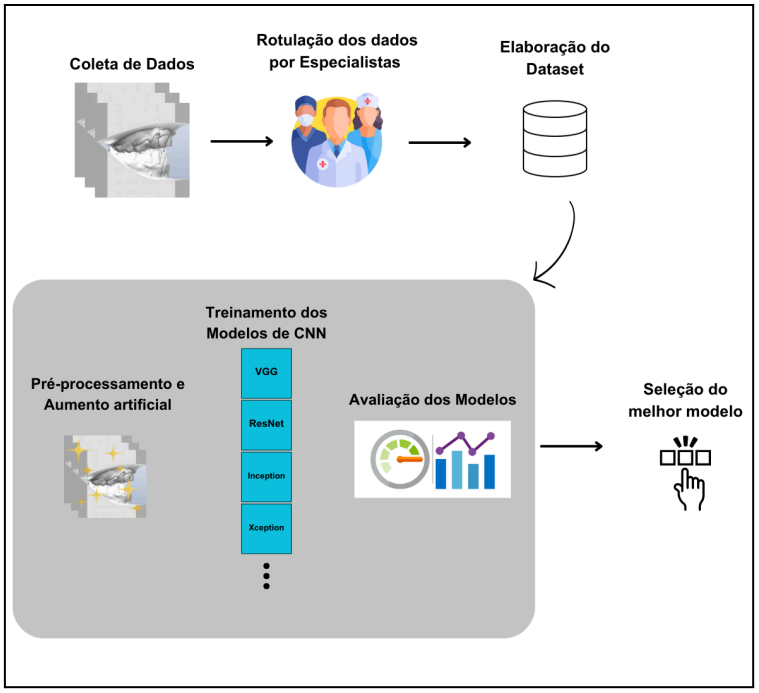
IMAGE-NET. Home Page. Disponível em: https://www.image-net.org/index.php. Acesso em: 1 abr. 2025.

A escassez de investigações focadas no tratamento ortodôntico de pacientes com fissuras labiopalatinas reforça a relevância e a originalidade deste estudo, que busca preencher essa lacuna, por meio do uso de redes neurais convolucionais para prever o prognóstico com base nos índices oclusais de Goslon. Embora este trabalho compartilhe semelhanças com o estudo de Huqh et al. (2023) acerca da predição e avaliação do crescimento maxilofacial, ele se diferencia pela utilização de imagens de escaneamentos intraorais, amplamente empregadas em ortodontia, em conjunto com o Índice de GOSLON.

3. Metodologia

Esta seção apresenta a metodologia utilizada para a condução deste estudo (Figura 2), incluindo a construção da base de dados, as técnicas de pré-processamento aplicadas nas imagens, as métricas de avaliação utilizadas, o ambiente de desenvolvimento e as arquiteturas de CNNs implementadas para a classificação das imagens de escaneamentos intraorais. Todas essas etapas serão descritas em detalhes nas próximas subseções.

Figura 2: Fluxograma da Metodologia.



Fonte: Autor.

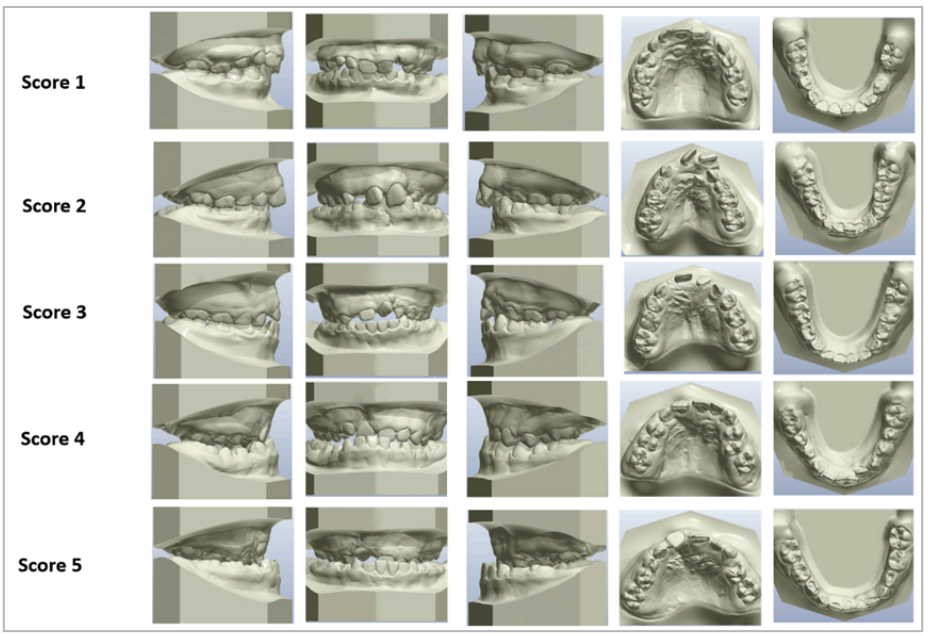
3.1. Base de Dados

Para este estudo, foram utilizadas imagens de escaneamentos intraorais dos arcos dentários superior e inferior. As imagens foram avaliadas por três especialistas em ortodontia com experiência no tratamento de fissuras labiopalatinas e rotuladas quanto aos índices oclusais de GOSLON e seu respectivo prognóstico. Posteriormente, os rótulos estabelecidos foram consolidados utilizando o coeficiente de Kappa de Cohen.

Os índices oclusais de GOSLON classificam os resultados de tratamento em excelente (escore 1), bom (escore 2), moderado (escore 3), pobre (escore 4) e muito pobre (escore 5) (Figura 3). Estes escores são atribuídos com base em três fatores clínicos: a relação anteroposterior, a relação vertical e a relação transversa. Eles foram agrupados segundo a categorização já utilizada pelos ortodontistas para o prognóstico de tratamento:

- Classe 1, com prognóstico favorável: escores 1 e 2;
- Classe 2, com prognóstico moderado: escores 3;
- Classe 3, com prognóstico desfavorável: escores 4 e 5.

Figura 3: Escaneamentos Intraorais e seus escores segundo o índice de GOSLON



Fonte: Imagens fornecidas pelo HULW.

A coleta de dados foi realizada no Serviço de Fissuras Labiopalatinas, do Hospital Universitário Lauro Wanderley (HULW), da Universidade Federal da Paraíba

(UFPB). O estudo é caracterizado como sendo do tipo transversal observacional e teve como universo todos os indivíduos cadastrados no Serviço de Fissuras Labiopalatinas entre os anos de 2015 e 2023[2], totalizando 2650 pacientes. Também fizeram parte do estudo todos os pacientes de ambos os sexos, com algum tipo de fissura palatina, com ou sem fissura labial associada, que foram atendidos no centro de fissuras no período da pesquisa, 1 de março de 2023 e 29 de fevereiro de 2024. Além desses, foram reutilizados dados em um estudo multicêntrico realizado sob os mesmos parâmetros entre os centros de tratamento: CADEFI-IMIP (Recife-PE), HRAC-USP (Bauru-SP) e Serviço de Fissuras Labiopalatinas do HULW-UFPB (João Pessoa-PB)

Foram incluídos na amostra os pacientes que concordaram em fazer parte da pesquisa e assinaram o Termo de Consentimento Livre e Esclarecido (TCLE) e o termo de assentimento. Os critérios de elegibilidade estabelecidos foram:

- Crianças com fissuras labiopalatinas não sindrômica com palato operado com pelo menos um ano de pós-operatório;
- Crianças que tenham modelos escaneados pré-tratamento ortodôntico, a partir dos 6 anos de idade;
- Foram excluídos indivíduos nascidos com fissuras labiopalatinas com idade acima de 20 anos, que apresentaram múltiplas perdas dentárias ou síndromes associadas, além dos casos considerados complexos pelos especialistas.

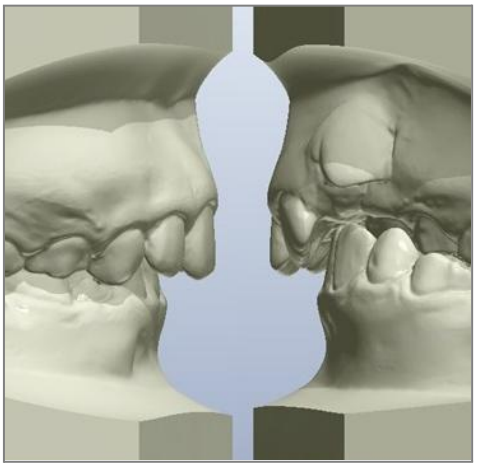
Ao todo foram coletados 587 escaneamentos intra orais, sendo 173 amostras pertencentes a Classe 1, 143 amostras pertencentes a Classe 2 e 271 amostras pertencentes a Classe 3. As imagens utilizadas foram produzidas a partir destes escaneamentos, capturando as vistas laterais e emparelhando-as frente a frente (Figura 4). Posteriormente foram salvas e organizadas em pastas, sendo uma para cada classe, em arquivo JPEG, com formato quadrado.

A divisão da base de dados foi realizada em duas etapas. Na primeira, aplicou-se a técnica de subamostragem (*undersampling*) com o objetivo de balancear a quantidade de dados em cada classe. Essa técnica reduz o número de amostras da classe majoritária, igualando-o ao da classe minoritária. Como resultado, cada classe passou a ter 143 amostras, totalizando 429 registros. Na segunda etapa, os dados foram divididos em

[2] No ano de 2015 foi criado o Grupo de Pesquisa em Anomalias Craniofaciais (GPAC), e foram instaurados protocolos de coleta de dados para fins de pesquisa, isto inclui também os escaneamentos intraorais, que são objetos deste estudo.

80%, destinados para treinamento e validação, e 20% para testes. A divisão foi feita de forma estratificada, mantendo a distribuição balanceada entre as classes em ambos os subconjuntos. Para garantir a reprodutibilidade dos experimentos, essa divisão foi salva em arquivos CSV contendo os nomes dos arquivos e seus respectivos rótulos.

Figura 4: Exemplo de imagem produzida a partir de um escaneamento.



Fonte: Autor.

3.2. Pré-processamento das Imagens

Após a coleta, foi realizado o pré-processamento das imagens. Essa etapa tem como objetivo garantir a qualidade dos dados, essencial para o processo de aprendizado de máquina. Para isso, foram utilizadas as ferramentas de manipulação de imagens do Keras e do TensorFlow.

Inicialmente, foram definidas as configurações padrão, que são comuns entre os subconjuntos de dados. Foi utilizado um *Data Loader* para gerar os lotes de imagens, contendo 32 amostras em cada *batch*. Também foi aplicada uma função de pré-processamento (cada rede fornece a sua própria função) que, em resumo, consiste na aplicação de normalização e transformações nos canais de cor. A aplicação desta função, garante que cada rede entenda os novos dados uma amostra similar ao conjunto de dados no qual foi pré-treinada.

Nas imagens utilizadas para o treinamento dos modelos foram aplicadas técnicas de aumento artificial de dados (*data augmentation*). Devido ao número reduzido de amostras, esse procedimento foi utilizado para expandir o subconjunto de dados, gerando imagens artificiais a partir das existentes. O *data augmentation* auxilia no

treinamento ao prevenir superajustamento (*overfitting*). Isto é feito sob a suposição de que mais informações podem ser extraídas do conjunto de dados original por meio de variações dos dados originais [23].

As transformações foram realizadas com o auxílio do módulo ImageDataGenerator da biblioteca Keras 2.15.0 (Tabela 1). Esse módulo atua aplicando transformações de forma estocástica a cada lote de imagens, com probabilidade de 0,5 para cada operação definida. Ou seja, a cada lote carregado, o algoritmo interno seleciona aleatoriamente quais transformações serão aplicadas a cada imagem, proporcionando diversidade sem comprometer a integridade do conteúdo. Para os subconjuntos de validação e teste, não foram aplicadas técnicas de aumento de dados, uma vez que essas imagens devem refletir fielmente a realidade do problema, possibilitando uma avaliação precisa do desempenho do modelo. Além disso, a semente aleatória para a geração dos lotes foi fixada com o valor 42, a fim de garantir a reprodutibilidade dos experimentos.

Tabela 1: Parâmetros para o módulo ImageDataGenerator

Parâmetro	Descrição	Valor
preprocessing_function	Normalização dos valores dos pixels	preprocess_input
rotation_range	Rotação aleatória aplicada dentro do intervalo especificado em graus.	15
width_shift_range	Translação horizontal aleatória, expressa como fração da largura da imagem.	0,1
height_shift_range	Translação vertical aleatória, expressa como fração da altura da imagem.	0,1
shear_range	Transformação de cisalhamento, que distorce a imagem ao mover uma parte dela mais do que outra.	0,1
zoom_range	Escalonamento aleatório, aumentando ou reduzindo a escala da imagem.	0,1
horizontal_flip	Reflexão horizontal da imagem para aumentar a diversidade dos dados.	True
brightness_range	Variação aleatória do brilho, multiplicando os valores dos pixels dentro do intervalo definido.	[0,8, 1,2]
fill_mode	Método de interpolação para preenchimento de áreas vazias geradas por transformações geométricas.	‘nearest'

Fonte: Autor.

3.3. Arquiteturas das CNNs

As CNNs são uma classe de redes neurais artificiais projetadas para processar dados estruturados em grades, como imagens [24]. Diferentemente das redes neurais tradicionais, as CNNs utilizam camadas convolucionais para extrair automaticamente padrões espaciais e hierárquicos dos dados de entrada [25]. Esse conceito foi desenvolvido no trabalho de Yann LeCun et al. (1989), combinando princípios de redes neurais artificiais com operações de convolução, para lidar com a alta dimensionalidade de imagens e outros tipos de dados espaciais.

Com o avanço das pesquisas em IA, especialmente na área da saúde, diferentes arquiteturas de CNN foram desenvolvidas para lidar com desafios específicos, como a extração mais eficaz de características e maior eficiência computacional. A escolha das arquiteturas para este estudo considerou tanto modelos amplamente utilizados na literatura, quanto redes mais recentes que apresentam melhorias em precisão e eficiência. Dentre elas, foram selecionadas as arquiteturas VGG16, VGG19, ResNet50, InceptionV3, Xception, DenseNet201, EfficientNetV2L e MobileNetV3L, cujas características serão discutidas a seguir e são resumidas na Tabela 2.

As arquiteturas VGG16 e VGG19 [27] são algumas das mais simples, consistindo basicamente em blocos convolucionais empilhados com filtros de tamanho 3×3. Os números 16 e 19 referem-se à quantidade de camadas em cada rede. Embora essas arquiteturas apresentem um bom desempenho em tarefas de reconhecimento de imagens, elas são computacionalmente exigentes devido ao grande número de parâmetros treináveis. A ResNet50 [28], por sua vez, se diferencia ao introduzir conexões residuais, que permitem uma melhor propagação dos gradientes e evitam o problema do desaparecimento do gradiente durante o treinamento.

As arquiteturas InceptionV3 [29] e Xception [30] compartilham conceitos semelhantes. A InceptionV3 utiliza módulos Inception, que combinam filtros convolucionais de diferentes tamanhos de *kernel*[3] em paralelo, permitindo a extração de informações em múltiplas escalas. Já a Xception aprimora essa abordagem ao empregar

[3]Em processamento de imagens e redes neurais convolucionais, *kernels* são pequenas matrizes que realizam a operação de convolução. Nesse processo, o *kernel* funciona como uma janela deslizante que percorre a imagem, aplicando operações matemáticas para extrair características. Em processamento de imagens, os valores dos *kernels* são pré-definidos para tarefas específicas, como detecção de bordas. Já em redes neurais convolucionais, esses valores são aprendidos automaticamente durante o treinamento, permitindo que o modelo identifique padrões relevantes para a tarefa.

convoluções separáveis em profundidade, reduzindo a complexidade computacional sem comprometer o desempenho. A DenseNet201 [31] se destaca pelo uso de conexões densas entre camadas, garantindo um fluxo mais eficiente de informações e um melhor aproveitamento dos recursos computacionais.

Tabela 2: Especificações das Arquiteturas Selecionadas e Acurácia desafio do ImageNet

Modelo	Tamanho(MB)	Parâmetros	Acurácia Top-1	Acurácia Top-5
Xception	88	22,9M	79,0%	94,5%
VGG16	528	138,4M	71,3%	90,1%
VGG19	549	143,7M	71,3%	90,0%
ResNet50	98	25,6M	74,9%	92,1%
InceptionV3	92	23,9M	77,9%	93,7%
MobileNetV3L	21	5,4M	75,0%	91,9%
DenseNet201	80	20,2M	77,3%	93,6%
EfficientNetV2L	479	119,0M	85,7%	97,5%

Fonte: https://keras.io/api/applications/. Acesso em: 11 abr. 2025

Arquiteturas mais recentes têm buscado equilibrar eficiência e desempenho, como é o caso da EfficientNetV2L [32] e da MobileNetV3L [33]. A EfficientNetV2L adota uma abordagem de escalonamento composto, ajustando simultaneamente profundidade, largura e resolução da rede de forma otimizada para maximizar a eficiência. Já a MobileNetV3L foi projetada para dispositivos móveis, considerando suas limitações computacionais. Para isso, ela emprega convoluções separáveis em profundidade, além de estratégias de redução progressiva do tamanho do *kernel* e funções de ativação mais eficientes.

A ampla variedade de arquiteturas de CNN reflete a constante evolução dessa tecnologia e permite avaliar o desempenho dessa abordagem em conjunto com diferentes técnicas projetadas para propósitos específicos. Além disso, essas arquiteturas foram amplamente testadas em desafios de reconhecimento de padrões em imagens e já

demonstraram sua eficácia em diversos estudos na área de IA aplicada à saúde, bem como nos trabalhos relacionados a este estudo.

3.4. Métricas de Avaliação

Para avaliar o desempenho dos modelos de CNN na tarefa proposta, foram utilizadas diversas métricas que permitem uma análise abrangente do comportamento do modelo. A escolha dessas métricas considerou a natureza do problema, caracterizado como uma classificação multiclasse, bem como a particularidade de ser uma aplicação na área da saúde. O processo de avaliação ocorre em duas etapas: a primeira durante o treinamento e a segunda na fase de testes.

Durante o treinamento, foram utilizadas as métricas de acurácia (1), que mede a proporção de predições corretas, e AUC (do inglês, *Area Under Curve*), Área Sob a Curva ROC (*Receiver Operating Characteristic*) [34,35], que fornece uma medida global do desempenho do modelo em diferentes limiares de decisão. A acurácia é calculada com base nos valores de Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN), onde:

- VP (Verdadeiro Positivo): instâncias corretamente classificadas na sua classe real;
- VN (Verdadeiro Negativo): instâncias que pertencem a outras classes e foram corretamente classificadas como não pertencentes à classe em questão;
- FP (Falso Positivo): instâncias de outra classe incorretamente classificadas como pertencentes a uma determinada classe;
- FN (Falso Negativo): instâncias de uma classe incorretamente classificadas como pertencentes a outra.

(1) $$Acurácia = \frac{VP+VN}{VP+VN+FP+FN}$$

Nesta etapa, as estatísticas são calculadas automaticamente pelo método *fit*, disponível para cada instância de modelo do Keras. Para facilitar a interpretação dessas métricas, foram gerados gráficos que ilustram a evolução do aprendizado do modelo ao longo das épocas. Também foram elaborados gráficos das curvas de erro, para identificação de comportamentos indesejados como *overfitting* e *underfitting* durante o treinamento e validação.

Na fase de testes, além da acurácia, foram empregadas as métricas precisão (2) [36], que mede a proporção de predições corretas entre as instâncias classificadas como positivas; revocação (3)[36], que indica a capacidade do modelo de identificar corretamente todas as instâncias de uma classe; e F1-*score* (4)[36], que combina precisão e revocação em uma média harmônica. Para essa etapa, as métricas foram calculadas utilizando as funções da biblioteca Scikit-Learn 1.6.1, complementadas por matrizes de confusão, permitindo uma avaliação detalhada do desempenho de cada modelo.

(2) $$Precisão = \frac{TP}{TP+FP}$$

(3) $$Revocação = \frac{TP}{TP+FN}$$

(4) $$F1 = 2 \cdot \frac{Precisão \cdot Revocação}{Precisão + Revocação}$$

3.5. Mapas de Calor

Uma das principais preocupações em aplicações de IA é a explicabilidade e interpretabilidade das previsões feitas pelos modelos, especialmente em contextos críticos como o da saúde. Com o avanço das CNNs, surgiram diversas abordagens para enfrentar esses desafios e aumentar a confiabilidade das previsões. Nesse cenário, um dos maiores avanços veio do desenvolvimento de técnicas baseadas na criação de mapas de ativação a partir dos gradientes, como o Grad-CAM, proposto por Selvaraju et al. (2017), que permite visualizar quais regiões da imagem mais influenciaram a decisão do modelo (Figura 4).

Para gerar os mapas de calor, a imagem é inicialmente processada pela CNN, percorrendo todas as suas camadas. Durante esse processamento, os gradientes são calculados em relação ao mapa de características de uma camada convolucional previamente escolhida, sendo comum a seleção da última camada por conter informações sobre o erro de previsão para a decisão final do modelo. O mapa de ativação, ou mapa de calor, é então obtido ponderando o mapa de características pelos valores médios globais dos gradientes. Por fim, o mapa gerado é redimensionado para coincidir com o tamanho da imagem original e sobreposto a ela, utilizando uma escala

de cores em que tons de vermelho indicam áreas de alta ativação e tons de azul representam regiões de menor influência na decisão do modelo [37].

Embora existam outras técnicas baseadas nos mesmos princípios do Grad-CAM, sua avaliação não é o foco deste estudo. Além de contribuir para a explicabilidade e interpretabilidade das previsões, essas técnicas de visualização também auxiliam na identificação e mitigação de possíveis vieses nos resultados, um fator especialmente relevante em aplicações sensíveis, como na área da saúde.

Figura 4: Grad-CAM utilizado para visualização da classe "cachorro"

Fonte: https://keras.io/examples/vision/grad_cam/. Acesso em: 25 mar. 2025.

3.6. Treinamento e Avaliação

O treinamento de cada arquitetura foi realizado no ambiente de desenvolvimento Google Collaboratory, utilizando GPUs T4 para acelerar o processo e garantir maior eficiência no treinamento dos modelos de CNN. O objetivo dessa etapa foi avaliar a capacidade de cada arquitetura diante do problema proposto. Para isso, foi adicionada uma camada de classificação padronizada, composta por uma camada totalmente conectada com 256 neurônios e função de ativação ReLU, seguida por uma camada de saída com 3 neurônios e função de ativação softmax.

Durante o treinamento, foram empregadas as técnicas *EarlyStopping*, *ReduceLROnPlateau* e *ModelCheckpoint*. Em resumo, essas técnicas monitoram o desempenho dos modelos ao longo do treinamento. O *EarlyStopping* observa a função de perda e interrompe o treinamento caso ela não apresente melhora após um determinado número de épocas, contribuindo para a prevenção de *overfitting*. O *ReduceLROnPlateau* reduz a taxa de aprendizado, quando uma métrica específica deixa de melhorar, ajudando o modelo a escapar de platôs e continuar evoluindo. Já o *ModelCheckpoint* salva os pesos do modelo durante o processo, garantindo que o

melhor estado obtido possa ser restaurado posteriormente. Esses recursos foram fundamentais para assegurar a estabilidade e a confiabilidade dos resultados.

Tabela 3: Configurações de parâmetros utilizados nos *callbacks*

Callback	**Parâmetros**
EarlyStopping	monitor='val_loss', patience=10, mode='auto'
ReduceLROnPlateau	monitor='val_accuracy', patience=5, factor=0.1, verbose=0, min_lr=0.000001
ModelCheckpoint	monitor='val_accuracy', verbose=1, save_best_only=True, mode='max'

Fonte: Autor

O otimizador Adam foi utilizado como padrão para todos os modelos, com uma taxa de aprendizado de 10^{-3}. Além disso, foi empregada a função de perda de entropia cruzada categórica, que assim como o otimizador Adam, é um parâmetro padrão do Keras para problemas de classificação com múltiplas classes.

Para garantir uma análise mais robusta do desempenho dos modelos foi utilizada uma técnica de validação cruzada, o *Stratified K-Fold*, da biblioteca Scikit-Learn 1.6.1. Este método foi empregado nos 80% separados para treinamento e validação, e consiste em dividir esse conjunto em K partes, de forma estratificada, ou seja, preservando a distribuição original das classes. A cada iteração, uma das partes é utilizada como conjunto de validação, enquanto as demais são usadas para o treinamento. Esse processo se repete até que cada uma das K partes tenha sido usada como validação uma vez.

As métricas de desempenho computadas em cada iteração foram agregadas utilizando a média, gerando as métricas da execução. Foram realizadas 10 execuções independentes com K=5, e ao final, foram calculadas as médias gerais e os respectivos desvios padrão de cada métrica, proporcionando uma estimativa mais confiável do desempenho do modelo. O *Stratified K-Fold* é comumente utilizado no contexto de aprendizado supervisionado, funcionando tanto para o ajuste de hiperparâmetros, quanto para a avaliação e comparação de desempenho entre modelos. Após a etapa de validação cruzada, a arquitetura com melhor desempenho médio foi selecionada como modelo preditivo final deste estudo.

Em seguida, esse modelo foi novamente treinado com a técnica de *fine-tuning*, com o objetivo de avaliar o seu desempenho no conjunto de testes. Para isso, realizou-se uma nova divisão a partir dos 80% anteriormente destinados para treinamento e

validação: 80% desse conjunto foram utilizados para o treino e 20% para validação, utilizada exclusivamente para o monitoramento dos *callbacks* (como *EarlyStopping* e *ModelCheckpoint*). Essa divisão também foi salva em arquivos CSV, contendo os nomes dos arquivos e seus respectivos rótulos, a fim de garantir a reprodutibilidade do experimento.

Por fim, o modelo foi avaliado no conjunto de testes, composto pelos 20% reservados no início do estudo. Os resultados de teste foram computados com a função classification_report, da biblioteca Scikit-Learn 1.6.1, que dispõe das métricas: acurácia, precisão, revocação e F1-*score*. Além disso, foram elaboradas matrizes de confusão para facilitar a interpretação dos resultados e fornecer uma visão mais detalhada do desempenho do modelo em cada classe.

4. Resultados

Nesta seção, apresentamos os resultados obtidos a partir da avaliação das diferentes arquiteturas de redes neurais convolucionais aplicadas ao problema proposto. O desempenho de cada rede é comparado utilizando as métricas estabelecidas na seção anterior, permitindo identificar a arquitetura mais adequada para esta aplicação. Além disso, posteriormente serão discutidas as implicações desses resultados no contexto da IA aplicada à Saúde, bem como suas possíveis limitações e direcionamentos para trabalhos futuros.

Ao término do treinamento de cada arquitetura, foi construída uma tabela contendo as médias e os desvios padrão de cada métrica coletada, calculados a partir das 10 repetições da validação cruzada com K=5. Essa tabela permite uma análise mais prática do desempenho de cada arquitetura, bem como da estabilidade dos modelos em diferentes cenários de treinamento, ou seja, para cada subconjunto da validação cruzada *K-fold*. A Tabela 4 apresenta as métricas de acurácia, precisão, revocação e F1-*score* para cada arquitetura de CNN avaliada.

De modo geral, as arquiteturas VGG apresentaram um desempenho superior em relação às demais, com acurácias de 83,3% e 83,5% para a VGG16 e a VGG19, respectivamente. Além disso, demonstraram consistência na classificação de imagens pertencentes a diferentes classes, o que pode ser observado nos valores muito próximos das métricas de F1-*score*, precisão e revocação, bem como nos escores de AUC superiores a 0,9. Esses resultados podem ser justificados pela estrutura dessas

arquiteturas que, apesar de simples, do ponto de vista tecnológico — compostas apenas por camadas convolucionais empilhadas com filtros de tamanho fixo —, são mais robustas em termos de quantidade de parâmetros, o que proporciona uma maior capacidade de capturar diferentes padrões nas imagens.

Tabela 4: Resultados de Avaliação das Arquiteturas

Arquitetura	Acurácia	+/-	Precisão	+/-	Revocação	+/-	F1-*score*	+/-	AUC	+/-
VGG19	**83,52%**	0,010	**0,84**	0,011	**0,83**	0,010	**0,84**	0,009	**0,93**	0,005
VGG16	**83,34%**	0,011	**0,84**	0,008	**0,83**	0,011	**0,83**	0,011	**0,93**	0,008
MobileNetV3L	**82,82%**	0,012	**0,84**	0,011	**0,83**	0,012	**0,83**	0,012	**0,93**	0,007
DenseNet201	**81,59%**	0,006	**0,83**	0,009	**0,82**	0,006	**0,82**	0,006	**0,92**	0,005
ResNet50	**79,59%**	0,009	**0,81**	0,007	**0,80**	0,009	**0,80**	0,010	**0,92**	0,007
InceptionV3	**74,57%**	0,019	**0,76**	0,020	**0,75**	0,019	**0,74**	0,021	**0,87**	0,010
Xception	**72,48%**	0,011	**0,74**	0,014	**0,72**	0,011	**0,72**	0,011	**0,85**	0,009
EfficientNetV2L	**72,01%**	0,011	**0,73**	0,011	**0,72**	0,011	**0,72**	0,012	**0,85**	0,013

Fonte: Autor.

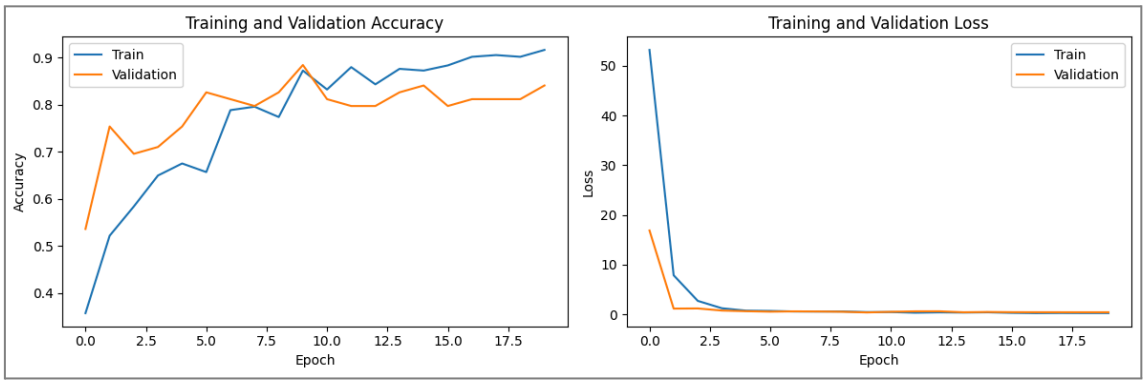
Em contraponto às VGG, que são naturalmente grandes em número de parâmetros, arquiteturas menores apresentaram desempenhos comparáveis. A MobileNetV3Large, por exemplo, alcançou 82,8% de acurácia e 0,83 de F1-*score*. Esse resultado demonstra que, para este problema, não há uma relação direta entre a quantidade de parâmetros e o desempenho obtido, uma vez que a MobileNetV3Large é a menor rede avaliada em termos de complexidade. Essa arquitetura foi especialmente desenvolvida para oferecer bom desempenho em dispositivos com baixo poder de processamento, o que a torna uma opção eficiente e viável para aplicações que exigem leveza e velocidade, sem grandes perdas de acurácia.

As arquiteturas DenseNet201, Xception, InceptionV3 e ResNet50 apresentaram resultados semelhantes. Entre elas, os melhores desempenhos foram obtidos pela rede DenseNet201, com 81,6% de acurácia e 0,82 de F1-*score*, e pela ResNet50, com 79,5% de acurácia e 0,80 de F1-*score*. O desempenho dessas arquiteturas pode ser atribuído às tecnologias que implementam, que permitem a captura de informações em diferentes

escalas (como nos blocos Inception, com múltiplos tamanhos de filtros convolucionais) e maior retenção de informações (por meio das conexões densas e conexões residuais).

Entre todas as arquiteturas avaliadas, a EfficientNetV2L obteve o pior desempenho, registrando 72% de acurácia e 0,72 de F1-*score*, além de 0,73 de precisão, 0,72 de revocação e 0,85 de AUC. Por ser a arquitetura com melhor desempenho no *benchmark* do ImageNet, era esperado que apresentasse resultados superiores. A quantidade reduzida de dados disponível pode ter sido um dos fatores que contribuíram para esse desempenho inferior. Além disso, a utilização de camadas de classificação fixas pode, de algum modo, ter limitado o potencial das redes. No entanto, o objetivo deste estudo é justamente avaliar qual arquitetura se adapta melhor ao problema proposto, utilizando o conjunto de dados como um *benchmark* específico, assim como é feito com o ImageNet.

Figura 5: Desempenho da VGG19 ao Longo das Épocas de Treinamento



Fonte: Autor.

Após a validação cruzada e comparações, a arquitetura VGG19 foi selecionada como modelo preditivo final para este estudo. Posteriormente ela foi novamente treinada, desta vez utilizando uma validação simples, apenas para direcionar os *callbacks* durante o treinamento. Durante o treinamento a rede apresentou um comportamento consideravelmente estável, atingindo valores de acurácia entre 80 e 90%, tanto no conjunto de treino, quanto no de validação (Figura 5). O processo de treinamento foi interrompido pelo EarlyStopping quando o valor da função de perda deixou de decair.

Após o treino, a rede foi avaliada no conjunto de testes e obteve 81% de acurácia e um F1-*score* de 0,81. Nos resultados da função classification_report (Tabela 5),

podemos destacar os valores: 0,71 de precisão e 0,76 de revocação para a Classe 2; 0,72 de revocação para a Classe 3; e 0,96 de revocação para a Classe 1. No geral, estes resultados indicam que o modelo encontrou dificuldades principalmente na distinção entre a Classe 2 e a Classe 3, entretanto, conseguiu captar bem as características da Classe 1. A confusão entre as Classes 2 e 3 pode ser justificada principalmente pelo fato de a Classe 2 compartilhar algumas características, em termos de relações interarcos, com a Classe 3.

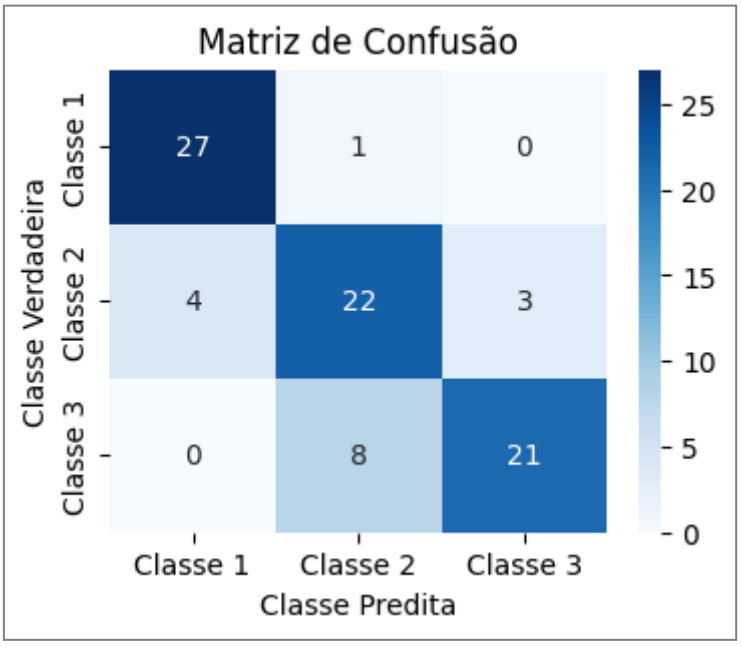
Tabela 5: Relatório de Classificação da VGG19 no Conjunto de Teste

	Precisão	**Revocação**	**F1-*score***	**Nº de Amostras**	**Acurácia**
Classe 1	**0,87**	**0,96**	**0,92**	**28**	**0,81**
Classe 2	**0,71**	**0,76**	**0,73**	**29**	
Classe 3	**0,88**	**0,72**	**0,79**	**29**	
Média	**0,82**	**0,81**	**0,81**	**86**	

Fonte: Autor.

Observando a matriz de confusão na Figura 6, nota-se que não houve, no conjunto de testes, confusão entre as Classes 1 e 3, o que, em termos de prognóstico e tratamento, é essencial, tendo em vista que, se um paciente com prognóstico de Classe 1 receber erroneamente um tratamento da Classe 3, isso pode acarretar em um tratamento equivocado e longo, bem como em danos estéticos e funcionais. Ademais, a confusão entre as Classes 2 e 3, em relação ao tratamento, também pode acarretar em danos ao paciente, a depender do tipo e do tempo da intervenção.

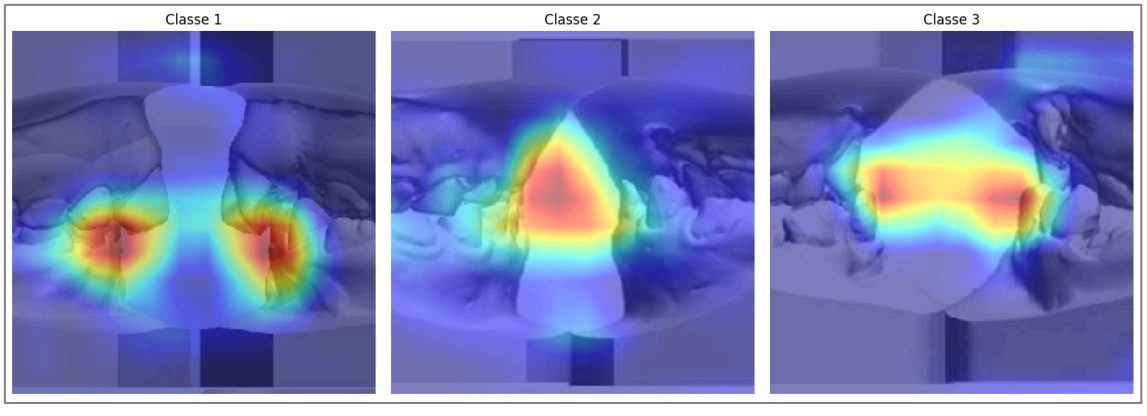
Figura 6: Matriz de confusão da VGG19 no conjunto de testes.



Fonte: Autor.

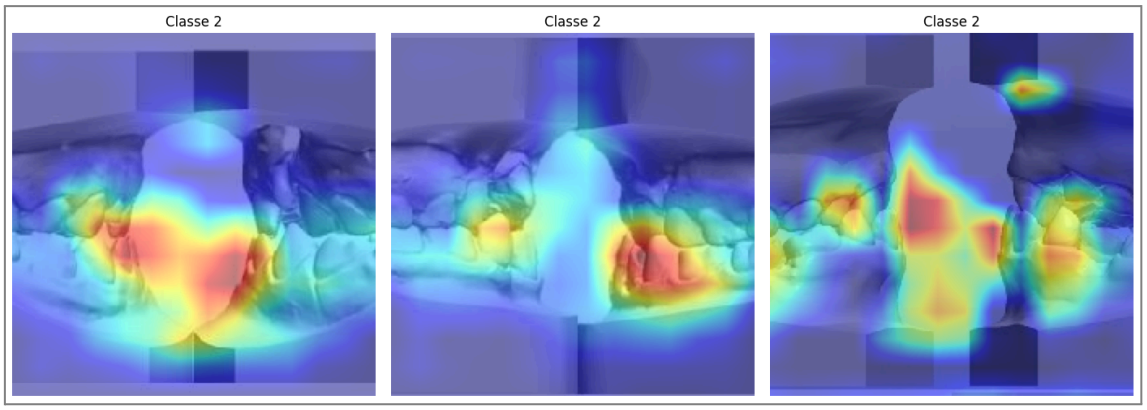
As visualizações geradas pelo Grad-CAM na Figura 7 destacam as regiões que o modelo considerou mais relevantes para a classificação. Em geral, os mapas de calor destacam bem os arcos dentais, que são a área de interesse. Além disso, as visualizações indicam que o modelo compreendeu, em algum nível, as relações horizontais entre os arcos. Para a Classe 1, onde o modelo obteve os melhores resultados, o mapa de calor parece destacar as duas laterais separadamente, e esse comportamento se manteve na maioria das imagens classificadas para esta classe. Enquanto isso, na maioria dos resultados para as Classes 2 e 3, os mapas de calor destacam as laterais em uma região de calor mais unida.

Figura 7: Visualizações Grad-CAM sobre previsões corretas no Conjunto de Teste



Fonte: Autor.

Figura 8: Visualizações Grad-CAM sobre previsões incorretas no Conjunto de Teste



Fonte: Autor.

A Figura 8 apresenta os mapas de calor correspondentes às previsões incorretas do modelo, especificamente nos casos em que os rótulos verdadeiros pertenciam à Classe 3, mas foram classificados como Classe 2. Sabe-se que os casos da Classe 3 (escores 4 e 5) envolvem um desalinhamento mais acentuado entre maxila e mandíbula quando comparados à Classe 2. Ao observar as amostras da Figura 8, nota-se que o grau de anomalia apresentado é consideravelmente menor em relação ao observado na amostra da Figura 7. Ainda assim, o Grad-CAM das duas primeiras imagens (da esquerda para a direita) evidencia, por meio das regiões em cores mais quentes, a atuação do modelo sobre os arcos inferiores, o que sugere que ele reconheceu a presença do desalinhamento, mas não conseguiu discriminar adequadamente o seu grau, levando à classificação incorreta entre as classes.

Dessa forma, observa-se que, entre todas as arquiteturas avaliadas, os modelos da família VGG se destacaram como os mais eficazes para o problema proposto. Apesar de sua simplicidade estrutural e alta complexidade em número de parâmetros, essas redes demonstraram um desempenho superior tanto em termos de acurácia, quanto de consistência nas métricas de classificação. Em especial, a VGG19 apresentou os melhores resultados gerais, consolidando-se como a arquitetura mais adequada para prever o prognóstico de crescimento maxilofacial dentro do conjunto analisado.

5. Conclusão

O presente estudo teve como objetivo avaliar o desempenho de diferentes arquiteturas de CNNs na classificação do prognóstico de crescimento maxilofacial em

pacientes com fissuras labiopalatinas, bem como propor um método de análise do prognóstico baseado em um modelo de IA.

Esse objetivo foi atingido, uma vez que foi possível realizar a avaliação das arquiteturas de forma adequada e em conformidade com as metodologias amplamente empregadas na literatura. Nesse sentido, destaca-se o desempenho geral satisfatório de todas as arquiteturas testadas, as quais apresentaram acurácia superior a 70%. Em especial, as arquiteturas da família VGG obtiveram resultados expressivos, com acurácia superior a 80%, sendo que a VGG-19 alcançou 83%. Vale mencionar, ainda, os modelos DenseNet201 e MobileNetV3Large, que também atingiram 80% de acurácia, evidenciando que, com o avanço das pesquisas em reconhecimento de imagens, é possível atingir desempenhos similares com menor custo computacional e temporal.

No que tange à interpretação das predições realizadas pelo modelo final sobre o conjunto de testes, a aplicação da técnica Grad-CAM demonstrou-se eficaz na identificação das regiões mais relevantes para a tomada de decisão do modelo. A interpretabilidade das saídas geradas constitui um aspecto fundamental no processo de adoção de tecnologias baseadas em IA, pois contribui para a geração de confiança e segurança em cada decisão automatizada — aspecto ainda mais crítico em áreas sensíveis como a saúde, cujo foco central é o bem-estar do paciente.

Por fim, este trabalho representa uma contribuição significativa para o tratamento de pacientes com fissuras labiopalatinas, destacando-se como o único, até o momento, com esse direcionamento específico. Os resultados obtidos ressaltam os benefícios da integração entre tecnologias emergentes, como a IA, e o campo da saúde, proporcionando maior suporte à tomada de decisão em uma área que carece de profissionais especializados.

5.1. Trabalhos Futuros

Embora os resultados obtidos neste estudo tenham se mostrado satisfatórios, é importante ressaltar que, dada a sensibilidade inerente à aplicação na área da saúde, tais resultados ainda não podem ser considerados ideais. Dessa forma, propõem-se, como trabalhos futuros, as seguintes direções de aprimoramento:

- Padronização na coleta de imagens: As imagens que compõem a base de dados deste estudo foram obtidas a partir de escaneamentos intraorais fornecidos por

diferentes centros clínicos, utilizando equipamentos distintos. Essa heterogeneidade na aquisição pode ter gerado distorções na distribuição das imagens, o que é especialmente crítico em bases de dados de dimensão reduzida. Uma possível abordagem, ainda não implementada nesta pesquisa, consiste na conversão das imagens para escala de cinza, com o intuito de eliminar variações cromáticas indesejadas e garantir maior uniformidade.

- Inclusão de informações clínicas relevantes: Em consonância com os achados de Huqh et al. (2023), no estudo intitulado "*Development of Artificial Neural Network-Based Prediction Model for Evaluation of Maxillary Arch Growth in Children with Complete Unilateral Cleft Lip and Palate*", variáveis como sexo e idade demonstraram impacto significativo na predição do crescimento maxilofacial. Assim, a incorporação dessas informações pode contribuir potencialmente para o aprimoramento do desempenho dos modelos preditivos.
- Incorporação de métricas específicas das relações interarcos: Considerando que o índice de Goslon classifica o prognóstico com base nas relações interarcos, a inclusão de dados quantitativos, como métricas de distância entre os arcos dentários, pode oferecer informações complementares relevantes ao modelo, aumentando sua capacidade discriminativa.

Adicionalmente, estão em andamento investigações quanto à viabilidade de utilização de outros formatos de imagem, como fotografias de modelos de gesso, ou da arcada dentária capturadas por meio de câmeras convencionais, ou de smartphones. Espera-se que, em um futuro próximo, o método desenvolvido possa ser integrado ao Sistema Único de Saúde (SUS), por meio de uma aplicação web ou aplicativo móvel, visando ampliar o acesso e aprimorar o tratamento de pacientes com fissuras labiopalatinas, especialmente em regiões remotas do Brasil, onde há escassez de profissionais especializados.

Por fim, o método proposto também apresenta potencial para aplicações voltadas ao controle de qualidade. Sua utilização pode viabilizar, por exemplo, a identificação de centros clínicos com maior incidência de prognósticos desfavoráveis (Classe 3), o que pode refletir falhas nos protocolos de tratamento adotados, ou deficiência na capacitação dos profissionais responsáveis.

6. Referências

1. **WORLD HEALTH ORGANIZATION (WHO)**. Oral Health. Genebra: WHO, mar. 2025. Disponível em: https://www.who.int/news-room/fact-sheets/detail/oral-health. Acesso em: 31 mar 2025.

2. MARAZITA, Mary L.; MOONEY, Mark P. Current concepts in the embryology and genetics of cleft lip and cleft palate. **Clinics in plastic surgery**, v. 31, n. 2, p. 125-140, 2004.

3. MOSSEY, Peter A. et al. Cleft lip and palate. **The Lancet**, v. 374, n. 9703, p. 1773-1785, 2009.

4. MARS, Michael et al. The Goslon Yardstick: a new system of assessing dental arch relationships in children with unilateral clefts of the lip and palate. **The Cleft palate journal**, v. 24, n. 4, p. 314-322, 1987.

5. CREPALDI, Jairo Lessa; ANDRÉ, Marcia; LOPEZ, Margareth Torrecillas. Análise da oclusão dentária em crianças portadoras de fissura labiopalatina bilateral. **Revista da Associação Paulista de Cirurgiões Dentistas**, v. 66, n. 4, p. 302-309, 2012.

6. YU, Kun-Hsing; BEAM, Andrew L.; KOHANE, Isaac S. Artificial intelligence in healthcare. **Nature biomedical engineering**, v. 2, n. 10, p. 719-731, 2018.

7. CANTU, Anselmo Garcia et al. Detecting caries lesions of different radiographic extension on bitewings using deep learning. **Journal of dentistry**, v. 100, p. 103425, 2020.

8. JÄRNSTEDT, Jorma et al. Comparison of deep learning segmentation and multigrader-annotated mandibular canals of multicenter CBCT scans. **Scientific Reports**, v. 12, n. 1, p. 18598, 2022.

9. WARIN, K. et al. Performance of deep convolutional neural network for classification and detection of oral potentially malignant disorders in

photographic images. **International Journal of Oral and Maxillofacial Surgery**, v. 51, n. 5, p. 699-704, 2022.

10. HUQH, Mohamed Zahoor Ul et al. Clinical applications of artificial intelligence and machine learning in children with cleft lip and palate—a systematic review. **International Journal of Environmental Research and Public Health**, v. 19, n. 17, p. 10860, 2022.

11. OROZCO-ARROYAVE, Juan Rafael et al. Automatic detection of hypernasal speech signals using nonlinear and entropy measurements. In: **INTERSPEECH**. 2012. p. 2029-2032.

12. MATHAD, Vikram C. et al. Deep learning based prediction of hypernasality for clinical applications. In: **ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).** IEEE, 2020. p. 6554-6558.

13. WANG, Xiyue et al. HypernasalityNet: Deep recurrent neural network for automatic hypernasality detection. **International Journal of Medical Informatics**, v. 129, p. 1-12, 2019.

14. GOLABBAKHSH, Marzieh et al. Automatic identification of hypernasality in normal and cleft lip and palate patients with acoustic analysis of speech. **The Journal of the Acoustical Society of America**, v. 141, n. 2, p. 929-935, 2017.

15. WANG, Xiyue et al. Automatic hypernasality detection in cleft palate speech using cnn. **Circuits, Systems, and Signal Processing**, v. 38, p. 3521-3547, 2019.

16. MACHADO, Renato Assis et al. Machine learning in prediction of genetic risk of nonsyndromic oral clefts in the Brazilian population. **Clinical Oral Investigations**, v. 25, p. 1273-1280, 2021.

17. ZHANG, Shi-Jian et al. Machine learning models for genetic risk assessment of infants with non-syndromic orofacial cleft. **Genomics, proteomics & bioinformatics**, v. 16, n. 5, p. 354-364, 2018.

18. LI, Yizhou et al. Clpnet: cleft lip and palate surgery support with deep learning. In: **2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).** IEEE, 2019. p. 3666-3672.

19. ALAM, Mohammad Khursheed et al. Sagittal jaw relationship of different types of cleft and non-cleft individuals. **Frontiers in pediatrics**, v. 9, p. 651951, 2021.

20. ALAM, Mohammad Khursheed; ALFAWZAN, Ahmed Ali. Dental characteristics of different types of cleft and non-cleft individuals. **Frontiers in Cell and Developmental Biology**, v. 8, p. 789, 2020.

21. SHAFI, Numan et al. Cleft prediction before birth using deep neural network. **Health informatics journal**, v. 26, n. 4, p. 2568-2585, 2020.

22. HUQH, Mohamed Zahoor Ul et al. Development of Artificial Neural Network-Based Prediction Model for Evaluation of Maxillary Arch Growth in Children with Complete Unilateral Cleft Lip and Palate. **Diagnostics**, v. 13, n. 19, p. 3025, 2023.

23. SHORTEN, Connor; KHOSHGOFTAAR, Taghi M. A survey on image data augmentation for deep learning. **Journal of big data**, v. 6, n. 1, p. 1-48, 2019.

24. GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge: MIT Press, 2016. Disponível em: https://mitpress.mit.edu/9780262035613/deep-learning/. Acesso em: 01 abr. 2025.

25. LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **nature**, v. 521, n. 7553, p. 436-444, 2015.

26. LECUN, Yann et al. Backpropagation applied to handwritten zip code recognition. **Neural computation**, v. 1, n. 4, p. 541-551, 1989.

27. SIMONYAN, Karen; ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv**:1409.1556, 2014.

28. HE, Kaiming et al. Deep residual learning for image recognition. In: Proceedings of the **IEEE conference on computer vision and pattern recognition**. 2016. p. 770-778.

29. SZEGEDY, Christian et al. Rethinking the inception architecture for computer vision. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. 2016. p. 2818-2826.

30. CHOLLET, François. Xception: Deep learning with depthwise separable convolutions. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. 2017. p. 1251-1258.

31. HUANG, Gao et al. Densely connected convolutional networks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. 2017. p. 4700-4708.

32. TAN, Mingxing; LE, Quoc. Efficientnet: Rethinking model scaling for convolutional neural networks. In: **International conference on machine learning**. PMLR, 2019. p. 6105-6114.

33. HOWARD, Andrew G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. **arXiv preprint arXiv**:1704.04861, 2017.

34. BRADLEY, Andrew P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. **Pattern recognition**, v. 30, n. 7, p. 1145-1159, 1997.

35. HUANG, Jin; LING, Charles X. Using AUC and accuracy in evaluating learning algorithms. **IEEE Transactions on knowledge and Data Engineering**, v. 17, n. 3, p. 299-310, 2005.

36. NAIDU, Gireen; ZUVA, Tranos; SIBANDA, Elias Mmbongeni. A review of evaluation metrics in machine learning algorithms. In: **Computer science on-line conference**. Cham: Springer International Publishing, 2023. p. 15-25.

37. SELVARAJU, Ramprasaath R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: **Proceedings of the IEEE international conference on computer vision**. 2017. p. 618-626.