

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Lucas Fiamoncini de Souza

Aprendizado profundo explicável com mecanismo de atenção para detecção de displasia epitelial oral em leucoplasias: Desenvolvimento de uma ferramenta educacional

Florianópolis
2026

Lucas Fiamoncini de Souza

Aprendizado profundo explicável com mecanismo de atenção para detecção de displasia epitelial oral em leucoplasias: Desenvolvimento de uma ferramenta educacional

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica para a obtenção do título de mestre em Engenharia Elétrica.
Orientador: Prof. Márcio Holsbach Costa, Dr.
Coorientador: Prof. Gilberto de Souza Melo, Dr.

Florianópolis
2026

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

de Souza, Lucas Fiamoncini
Aprendizado profundo explicável com mecanismo de atenção para detecção de displasia epitelial oral em leucoplasias: Desenvolvimento de uma ferramenta educacional / Lucas Fiamoncini de Souza ; orientador, Márcio Holsbach Costa, coorientador, Gilberto de Souza Melo, 2026.
89 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, 2026.

Inclui referências.

1. Engenharia Elétrica. 2. Displasia epitelial oral. 3. Aprendizado profundo. 4. Mecanismo de Atenção. 5. Educação em patologia bucal. I. Costa, Márcio Holsbach. II. Melo, Gilberto de Souza. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. IV. Título.

Lucas Fiamoncini de Souza

Aprendizado profundo explicável com mecanismo de atenção para detecção de displasia epitelial oral em leucoplasias: Desenvolvimento de uma ferramenta educacional

O presente trabalho em nível de Mestrado foi avaliado e aprovado, em 19 de junho de 2026, pela banca examinadora composta pelos seguintes membros:

Prof. Bruno Catarino Bispo, Dr.
Universidade Federal de Santa Catarina

Prof. Danilo Silva, Ph.D
Universidade Federal de Santa Catarina

Prof. Paulo Rogério de Faria, Dr.
Universidade Federal de Uberlândia

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Mestre em Engenharia Elétrica.



Documento assinado digitalmente
Carlos Renato Rambo
Data: 08/07/2026 15:44:27-0300
CPF: ***.748.198-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Carlos Renato Rambo, Dr
Coordenador do Programa



Documento assinado digitalmente
Marcio Holsbach Costa
Data: 08/07/2026 09:55:05-0300
CPF: ***.170.680-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Márcio Holsbach Costa, Dr.
Orientador

Florianópolis, 2026.

AGRADECIMENTOS

Agradeço à minha família pelo total apoio e incentivo durante toda a minha jornada acadêmica. Agradeço aos meus orientadores, professor Márcio Holsbach Costa e professor Gilberto de Souza Melo, pelas orientações, paciência e apoio incondicional ao longo deste trabalho. À professora Elena Riet Correa Rivero pelo fornecimento do banco de dados e por suas contribuições. À UFSC por fornecer um ambiente acadêmico estimulante e recursos essenciais para a realização deste trabalho. Por fim, agradeço a todos que, de alguma forma, contribuíram para a realização deste trabalho e minha formação acadêmica.

RESUMO

O câncer de boca é uma das neoplasias malignas mais prevalentes, frequentemente precedido por lesões orais potencialmente malignas, como a leucoplasia. A displasia epitelial oral (OED) é uma condição histopatológica de risco aumentado para transformação maligna, cuja detecção e classificação em graus de severidade apresenta elevada subjetividade, resultando em inconsistências diagnósticas. Nos últimos anos, técnicas de inteligência artificial têm sido exploradas para auxiliar na detecção e classificação da OED, embora a falta de interpretabilidade dos modelos e a escassez de aplicações educacionais permaneçam como lacunas significativas na literatura. Nesse contexto, este trabalho propõe o desenvolvimento de uma plataforma educacional baseada em aprendizado profundo explicável com mecanismo de atenção para auxiliar estudantes de pós-graduação em patologia bucal. O sistema integra imagens histopatológicas e dados clínicos de pacientes diagnosticados com leucoplasia para detecção de OED. Técnicas de explicabilidade baseadas em mapas de calor são empregadas para destacar as regiões histológicas mais relevantes para a decisão do modelo, sendo essas regiões validadas por comparação sistemática com anotações de patologistas especialistas. Os resultados obtidos são equiparáveis ao estado da arte, com F1-score de 0,92 e AUC de 0,94. Um estudo de caso com estudantes de pós-graduação demonstrou que as informações de apoio fornecidas pela ferramenta aumentaram a acurácia do processo de classificação de 0,63 $\pm$ 0,09 para 0,82 $\pm$ 0,14 ($p < 0,01$) e a especificidade de 0,46 $\pm$ 0,22 para 0,78 $\pm$ 0,26 ($p < 0,01$) — sugerindo seu potencial pedagógico. A plataforma desenvolvida integra predições, valores de confiabilidade e mapas de atenção, promovendo um aprendizado ativo e visualmente orientado em patologia bucal.

Palavras-chave: Displasia epitelial oral. Inteligência artificial. Aprendizado profundo. Mecanismo de atenção. Explicabilidade. Educação em patologia bucal.

ABSTRACT

Oral cancer is one of the most prevalent malignant neoplasms, frequently preceded by potentially malignant oral lesions such as leukoplakia. Oral epithelial dysplasia (OED) is a histopathological condition with an increased risk of malignant transformation, whose detection and classification into severity grades presents high subjectivity, resulting in diagnostic inconsistencies. In recent years, artificial intelligence techniques have been explored to assist in the detection and classification of OED, although the lack of interpretability of the models and the scarcity of educational applications remain significant gaps in the literature. In this context, this work proposes the development of an educational platform based on explainable deep learning with attention mechanism to assist postgraduate students in oral pathology. The system integrates histopathological images and clinical data from patients diagnosed with leukoplakia for the detection of OED at the clinical case level. Explainability techniques based on heatmaps are employed to highlight the most relevant histological regions for the model's decision, these regions being validated by systematic comparison with annotations from expert pathologists. The results obtained are comparable to the state of the art, with an F1-score of 0.92 and an AUC of 0.94. A case study with postgraduate students demonstrated that the supporting information provided by the tool increased the accuracy of the classification process from 0.63 ± 0.09 to 0.82 ± 0.14 ($p < 0.01$) and the specificity from 0.46 ± 0.22 to 0.78 ± 0.26 ($p < 0.01$) — suggesting its pedagogical potential. The developed platform integrates predictions, confidence scores, and attention maps, promoting active and visually oriented learning in oral pathology.

Keywords: Oral epithelial dysplasia. Artificial intelligence. Deep learning. Attention mechanism. Explainability. Education in oral pathology.

LISTA DE FIGURAS

Figura 1 – Exemplo de leucoplasia. 26

Figura 2 – Exemplo de uma fotografia com ampliação de 200 vezes de uma lâmina histológica corada com hematoxilina e eosina referente a uma lesão bucal; em que "A" indica a camada de queratina, "B" o tecido epitelial e "C" o tecido conjuntivo. 26

Figura 3 – Evolução da displasia epitelial. 27

Figura 4 – Exemplo de dilatação, em que (a) representa uma imagem I composta de um objeto A e seu fundo (*background*); (b) é o elemento estruturante B; e (c) é o resultado da dilatação de A por B, denotado por $A \oplus B$. 28

Figura 5 – Representação do SVM de margens suaves. 31

Figura 6 – Exemplo de uma imagem 3×3 convoluída com um filtro 2×2, resultando em um mapa 2×2. 34

Figura 7 – Ilustração da operação de *max pooling*, no qual blocos de 2×2 *pixels* em um mapa de atributos são combinados usando o operador *max* para gerar um novo mapa de atributos de menor dimensionalidade. 35

Figura 8 – Exemplo de uma rede residual com três blocos residuais. 36

Figura 9 – Exemplo de implementação do método CAM. 37

Figura 10 – Exemplo de implementação do algoritmo Grad-CAM. 38

Figura 11 – Exemplo de implementação da arquitetura ABN. 39

Figura 12 – Matriz de confusão para um problema de classificação binária. 41

Figura 13 – Curva ROC (linha contínua azul) para um problema de classificação binária. Em laranja pontilhado é apresentado o resultado de um classificador aleatório. 42

Figura 14 – Exemplo de mapa de atenção: (a) Exemplo de imagem histopatológica; (b) Mapa de atenção da imagem histopatológica. 48

Figura 15 – Diagrama em blocos da metodologia proposta para detecção de displasia epitelial oral utilizando o modelo base. 49

Figura 16 – Diagrama em blocos da metodologia proposta para detecção de displasia epitelial oral utilizando o modelo ABN. 50

Figura 17 – Esquema de utilização da plataforma educacional para detecção de displasia epitelial oral. 53

Figura 18 – Distribuição das métricas de desempenho em gráficos de violino: (a) modelo base e (b) ABN. 56

Figura 19 – Matriz de confusão para: (a) modelo base e (b) ABN. 56

Figura 20 – Curva ROC e medida AUC para: (a) modelo base e (b) ABN. 57

Figura 21 – Exemplos de imagem histológica e mapa de atenção provenientes do modelo base pertencentes a casos: (a-b) corretamente classificado como displásico com percentual de confiabilidade de 100%; (c-d) corretamente classificado como displásico com percentual de confiabilidade de 99%; (e-f) erroneamente classificado como não displásico com percentual de confiabilidade de 67%; (g-h) erroneamente classificado como displásico com percentual de confiabilidade de 52%. 59
Figura 22 – Histograma de incremento percentual da área anotada após a aplicação da dilatação morfológica. 60
Figura 23 – Resultados da métrica de Precisão Top-K% para diferentes valores de K, para a ROI original e com dilatação para o modelo base. 60
Figura 24 – Resultados da métrica de Fração de Ativação para a ROI original e com dilatação para o modelo base. 61
Figura 25 – Resultados da métrica *Pointing Game* para a ROI original e com dilatação para o modelo base. 62
Figura 26 – Exemplos de imagem histológica e mapa de atenção provenientes do modelo ABN pertencentes a casos: (a-b) corretamente classificado como displásico com percentual de confiabilidade de 100%; (c-d) corretamente classificado como displásico com percentual de confiabilidade de 100%; (e-f) erroneamente classificado como não displásico com percentual de confiabilidade de 100%; (g-h) erroneamente classificado como não displásico com percentual de confiabilidade de 100%. 63
Figura 27 – Resultados da métrica de Precisão Top-K% para diferentes valores de K, para a ROI original e com dilatação para o modelo ABN. 64
Figura 28 – Resultados da métrica de Fração de Ativação para a ROI original e com dilatação para o modelo ABN. 65
Figura 29 – Resultados da métrica *Pointing Game* para a ROI original e com dilatação para o modelo ABN. 65
Figura 30 – Ferramenta educacional desenvolvida para auxílio na detecção de displasia epitelial oral. 68
Figura 31 – Seção da plataforma para inserção de dados complementares do paciente. . . 68
Figura 32 – Resultado da acurácia dos participantes para etapa 1 (sem informações de apoio) e etapa 2 (com informações de apoio). 71
Figura 33 – Resultado da sensibilidade dos participantes para etapa 1 (sem informações de apoio) e etapa 2 (com informações de apoio). 71
Figura 34 – Resultado da especificidade dos participantes para etapa 1 (sem informações de apoio) e etapa 2 (com informações de apoio). 72
Figura 35 – Resultado do questionário de avaliação das informações de apoio. 73

LISTA DE TABELAS

Tabela 1 – Faixa de valores dos hiperparâmetros avaliados durante o processo de *grid search* para a ResNet50. 49
Tabela 2 – Faixa de valores dos hiperparâmetros avaliados durante o processo de *grid search* para a ABN. 51
Tabela 3 – Faixa de valores dos hiperparâmetros avaliados durante o processo de *grid search* para o SVM. 51
Tabela 4 – Valores médios das métricas de desempenho, desvio padrão e intervalo de confiança de 95% para o modelo base e o modelo ABN. 56
Tabela 5 – Comparação das métricas entre as máscaras binárias originais e dilatadas para o modelo base. 61
Tabela 6 – Comparação das métricas entre as máscaras binárias originais e dilatadas para o modelo ABN. 66
Tabela 7 – Métricas de desempenho dos participantes sem e com as informações de apoio. 72
Tabela 8 – Acurácia dos participantes para cada classe sem e com as informações de apoio. 72

LISTA DE ABREVIATURAS E SIGLAS

ABN	*Attention Branch Networks*
AUC	*Area Under Curve*
CAM	*Class Activation Mapping*
CNN	*Convolutional Neural Network*
FC	*Fully Connected*
FN	Falso Negativo
FP	Falso Positivo
GAP	*Global Average Pooling*
Grad-CAM	*Gradient-Weighted Class Activation Mapping*
IA	Inteligência Artificial
OED	*Oral Epithelial Dysplasia*
OMS	Organização Mundial da Saúde
RBF	*Radial Basis Function*
ReLU	*Rectified Linear Unit*
ROC	*Receiver Operating Characteristic*
ROI	*Region of Interest*
SVM	*Support Vector Machine*
TFP	Taxa de Falsos Positivos
TVP	Taxa de Verdadeiros Positivos
ViT	*Vision Transformer*
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

SUMÁRIO

1 INTRODUÇÃO 19
1.1 TRABALHOS RELACIONADOS 20
1.2 OBJETIVOS 22
1.2.1 Objetivos específicos 22
1.3 ORGANIZAÇÃO DO TRABALHO 23

2 FUNDAMENTAÇÃO TEÓRICA 25
2.1 DISPLASIA EPITELIAL ORAL 25
2.2 PROCESSAMENTO DE IMAGENS DIGITAIS 25
2.2.1 Morfologia matemática 27
2.2.1.1 Dilatação 28
2.3 APRENDIZADO DE MÁQUINA E APRENDIZADO PROFUNDO 28
2.3.1 Máquina de vetores de suporte 29
2.3.2 Redes neurais artificiais 30
2.3.3 Redes convolucionais 33
2.3.3.1 Camadas convolucionais e de ativação 33
2.3.3.2 Camadas de pooling 34
2.3.3.3 Camadas de saída 34
2.3.4 ResNet 35
2.4 MECANISMOS DE ATENÇÃO E INTERPRETABILIDADE EM REDES CONVOLUCIONAIS 36
2.4.1 Método CAM 36
2.4.2 Método Grad-CAM 37
2.4.3 Attention Branch Networks 38
2.4.3.1 Extrator de atributos 38
2.4.3.2 Ramo de atenção 39
2.4.3.3 Mecanismo de atenção 39
2.4.3.4 Ramo de percepção 40
2.4.3.5 Função de perda composta 40
2.5 AVALIAÇÃO DE DESEMPENHO 40
2.5.1 Matriz de confusão 40
2.5.2 Métricas de desempenho 41
2.5.3 Métricas de avaliação dos mapas de atenção 42

3 METODOLOGIA 45
3.1 BANCO DE DADOS 45
3.1.1 Banco de imagens 45
3.1.2 Dados complementares dos pacientes 46

3.1.3 Redimensionamento de imagens e separação de conjuntos de dados . . . 46
3.2 ANOTAÇÃO DAS REGIÕES DE INTERESSE E GERAÇÃO DOS MAPAS DE ATENÇÃO DE REFERÊNCIA . . . 47
3.3 METODOLOGIA DE DETECÇÃO DE DISPLASIA EPITELIAL ORAL . 47
3.3.1 Modelo base . . . 48
3.3.2 Modelo ABN . . . 49
3.3.3 Classificador SVM . . . 51
3.4 ANÁLISE COMPARATIVA DOS MAPAS DE ATENÇÃO . . . 51
3.5 METODOLOGIA DE IMPLEMENTAÇÃO DA PLATAFORMA EDUCACIONAL . . . 52
3.5.1 Metodologia de avaliação por estudantes de pós-graduação em patologia bucal . . . 52

4 RESULTADOS . . . 55
4.1 DESEMPENHO DE CLASSIFICAÇÃO . . . 55
4.1.1 Influência da integração dos dados complementares . . . 55
4.2 DESEMPENHO DOS MAPAS DE ATENÇÃO . . . 57
4.2.1 Modelo base . . . 58
4.2.2 Modelo ABN . . . 62
4.3 PLATAFORMA EDUCACIONAL . . . 66
4.4 AVALIAÇÃO DAS INFORMAÇÕES DE APOIO DA PLATAFORMA . . . 69

5 CONCLUSÃO . . . 75

REFERÊNCIAS . . . 77

APÊNDICE A – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO . . . 81

APÊNDICE B – FOLHA DE AVALIAÇÃO APLICADA NA ETAPA 1 DA PESQUISA . . . 85

APÊNDICE C – QUESTIONÁRIO APLICADO NA ETAPA 2 SOBRE AS INFORMAÇÕES DE APOIO FORNECIDAS PELO MODELO . . . 87

1 INTRODUÇÃO

O câncer de boca é um dos tipos mais comuns de neoplasias malignas, sendo o tabagismo e o etilismo alguns dos principais fatores etiológicos (Khanagar et al., 2023). De modo geral, o câncer de boca é precedido por um conjunto de alterações da mucosa oral, que se apresentam clinicamente como diferentes lesões, tais como a leucoplasia e eritroplasia, sendo classificadas como lesões orais potencialmente malignas (Odell et al., 2021). O diagnóstico precoce dessas lesões é imprescindível, sendo usualmente realizado por meio de biópsia e análise histopatológica. Nesse contexto, essas lesões podem apresentar diferentes características histopatológicas, tais como hiperceratose, acantose e displasia epitelial (Warnakulasuriya et al., 2020).

A displasia epitelial oral (OED — do inglês, *Oral Epithelial Dysplasia*) refere-se a um conjunto de alterações arquiteturais e citológicas do epitélio, relacionadas a um acúmulo de mutações genéticas e associadas a um risco aumentado de transformação maligna (Hankinson et al., 2024). A detecção precoce dessa condição é essencial para o sucesso do tratamento e diminuição das taxas de transformação maligna e de mortalidade (Silva et al., 2024). Conforme critérios da Organização Mundial da Saúde (OMS), a OED pode ser classificada em difentes graus de severidade, de acordo com os terços do epitélio acometidos pelas alterações displásicas (Muller; Tilakaratne, 2022). No entanto, a identificação e a classificação da OED em diferentes níveis de severidade é altamente subjetiva, o que pode resultar em inconsistência diagnóstica entre os patologistas (Alajaji et al., 2024).

Nos últimos anos, diversos estudos têm investigado a aplicação de técnicas de inteligência artificial (IA) na detecção e classificação de OED (Alajaji et al., 2024). Apesar dos avanços alcançados, a ausência de mecanismos de explicabilidade — sobretudo em modelos baseados em aprendizado profundo — permanece como uma limitação significativa e um campo em aberto para investigação. Além disso, a maior parte dos estudos com IA em patologia bucal está centrada no suporte ao diagnóstico clínico, sendo pouco explorado seu potencial no contexto educacional (Abdul et al., 2024). Observa-se, assim, uma lacuna importante no aproveitamento dessas tecnologias como instrumentos pedagógicos.

O ensino de patologia bucal, tradicionalmente baseado em aulas expositivas e observação ao microscópio, enfrenta desafios relacionados à subjetividade na interpretação das lâminas histológicas e ao engajamento discente diante da complexidade das alterações teciduais. Nesse contexto, poucos trabalhos têm explorado a utilização de ferramentas educacionais na área da patologia, embora resultados promissores em termos de desempenho acadêmico e engajamento discente tenham sido observados (Kanthan; Senger, 2011; Hassell et al., 2022). Além disso, a utilização de IA em ambientes educacionais tem se mostrado promissora, apresentando benefícios tanto a discentes quanto a docentes (Chen; Chen; Lin, 2020). Sendo assim, ferramentas educacionais impulsionadas por IA representam uma oportunidade inovadora para aprimorar a formação de estudantes na área da saúde. A utilização de algoritmos de explicabilidade — como o Grad-CAM (*Gradient-Weighted Class Activation Mapping*) (Selvaraju et al., 2017) —

pode, por exemplo, destacar as regiões mais relevantes das imagens para a decisão do modelo, proporcionando aos estudantes uma visão mais clara das alterações arquiteturais e citológicas associadas à presença de displasia. Além disso, interfaces interativas que permitam a navegação em imagens histológicas digitalizadas com apoio de modelos de IA podem promover um aprendizado ativo, personalizado e visualmente orientado.

Transformar soluções desenvolvidas inicialmente para diagnóstico em ambientes de aprendizado pode facilitar a compreensão de conceitos complexos e aumentar o interesse dos estudantes por disciplinas fundamentais como a patologia bucal. A construção de plataformas educacionais baseadas em IA também pode democratizar o acesso a conteúdos de alta qualidade, especialmente em contextos com escassez de profissionais especializados ou infraestrutura laboratorial limitada. Assim, a proposta de desenvolvimento de ferramentas educacionais com suporte de inteligência artificial não apenas acompanha o avanço tecnológico da área médica, mas também responde a demandas pedagógicas contemporâneas ao integrar inovação, explicabilidade e engajamento discente no processo de ensino em patologia bucal.

1.1 TRABALHOS RELACIONADOS

O trabalho de Alajaji et al. (2024) apresenta uma revisão sobre o uso de IA na detecção e classificação em graus de severidade de OED. Nesse trabalho, foram analisados artigos datados de 1980 a janeiro de 2024. Após análise criteriosa, 24 artigos foram selecionados para integrar a revisão, sendo 19 deles focados em aprendizado profundo e 4 em aprendizado clássico baseado em extração de atributos. Os autores ressaltam que, apesar dos avanços de desempenho obtidos, a falta de transparência e interpretabilidade continua sendo um desafio significativo. Eles sugerem que a integração de técnicas de explicabilidade pode melhorar a confiança dos profissionais de saúde nos sistemas baseados em IA. Nesse cenário, o trabalho de Afify, Mohammed e Hassanien (2023) destaca-se ao empregar o método Grad-CAM como ferramenta de interpretação dos modelos preditivos, permitindo a visualização das regiões mais relevantes das imagens histopatológicas para a detecção do carcinoma de células escamosas; estágio avançado de câncer oral. A utilização do Grad-CAM em detecção de carcinoma de células escamosas também é observado em Saldivia-Siracusa et al. (2025), trabalho que se destaca ao avaliar por meio de especialistas a confiabilidade e localização focal dos mapas gerados.

Os autores de Mitsuhara et al. (2019) propõem uma abordagem para incorporar o conhecimento humano em redes neurais profundas por meio de mapas de atenção através da arquitetura de redes de ramos de atenção (ABN — do inglês, *Attention Branch Networks*). A metodologia apresentada visa melhorar o desempenho e interpretabilidade dos modelos, permitindo que eles se concentrem em regiões relevantes das imagens, conforme identificado por especialistas humanos. Nesse contexto, vale destacar que a qualidade das anotações realizadas por especialistas é determinante para o desempenho dos modelos, uma vez que a subjetividade inerente à interpretação humana é transferida ao modelo, que passa a quantificar e reproduzir a percepção histopatológica do profissional, tornando-se uma extensão do seu raciocínio diag-

nóstico (Araújo et al., 2023). Embora o estudo desenvolvido em Mitsuhara et al. (2019) não seja especificamente focado em patologia bucal, a abordagem proposta pode ser adaptada para melhorar a explicabilidade dos modelos de aprendizado profundo aplicados à detecção de OED. Isso pode ser verificado no trabalho de Song et al. (2023), em que os autores utilizam uma abordagem muito similar aplicada à detecção de câncer bucal; com a arquitetura ABN apresentando melhor desempenho em comparação com outras arquiteturas analisadas.

Em Peng et al. (2024), é proposto um sistema de aprendizado profundo para a classificação de OED em diferentes graus de severidade com base nos critérios estabelecidos pela OMS. O modelo desenvolvido gera mapas de calor que destacam as regiões da imagem histológica associadas a cada grau de displasia, contribuindo assim para a interpretabilidade do processo de decisão do algoritmo. No entanto, o estudo apresenta limitações importantes. Primeiramente, não há uma análise comparativa entre os mapas de calor gerados pelo modelo e a interpretação fornecida por patologistas especialistas, o que compromete a validação da explicabilidade da abordagem proposta. Além disso, os autores utilizam o rótulo atribuído à imagem histológica completa como referência para os *patches* (recortes) extraídos da mesma imagem. Essa estratégia pode introduzir ruído nos dados, uma vez que diferentes regiões de uma única lâmina podem apresentar distintos graus de displasia, comprometendo a consistência dos rótulos atribuídos aos *patches*. O melhor resultado obtido pelos autores foi com o modelo EfficientNet-B0 (Tan; Le, 2019), apresentando uma acurácia de 81,3% e uma área sob a curva (AUC — do inglês *Area Under the Curve*) de 0,793.

Uma metodologia que utiliza dados clínicos e imagens histológicas de pacientes clinicamente diagnosticados com leucoplasia foi proposta por Rocha et al. (2025). O trabalho mostrou que a combinação de dados clínicos e imagens histológicas pode melhorar o desempenho dos modelos de aprendizado de máquina para detecção da OED proveniente de casos de leucoplasias; servindo como motivação do presente trabalho. Outra grande contribuição deste trabalho foi a detecção da displasia a nível de casos, não somente a nível de imagens histológicas, proporcionando uma abordagem mais alinhada com as práticas estabelecidas entre patologistas orais. Nesse trabalho, o melhor desempenho obtido foi utilizando uma ResNet50 (He et al., 2015), com F1-*score* de 0,91 e AUC de 0,95 na detecção de displasia. Entretanto, uma limitação de Rocha et al. (2025) é a necessidade de segmentação manual das imagens histológicas, o que pode ser um processo demorado e sujeito a erros. Além disso, o trabalho carece de uma análise mais aprofundada sobre a explicabilidade dos modelos de aprendizado de máquina utilizados.

Os trabalhos de Pedroso et al. (2026) e Pontes et al. (2025) também incorporam dados clínicos e imagens histológicas em abordagens de aprendizado de máquina, neste caso voltados à predição de risco de malignidade em leucoplasias, alcançando AUCs de 0,77 e 0,85, respectivamente. O trabalho de Pontes et al. (2025) vai além ao incluir conteúdo de DNA como atributo adicional.

O trabalho de Hadilou et al. (2025) se destaca por apresentar pela primeira vez uma abordagem utilizando *Vision Transformers* (ViT) para a classificação de graus de severidade de

OED. Os autores apresentam os resultados para dois tipos de classificação: (i) 94% de acurácia para classificação em três classes (tecido epitelial normal, baixo risco de malignidade e alto risco de malignidade); e (ii) 97% de acurácia para classificação em quatro classes (tecido epitelial normal, displasia leve, displasia moderada e displasia severa). Além dos resultados promissores, o trabalho também apresenta os mapas de atenção gerados pelo modelo ViT. Entretanto, os autores não realizam uma análise comparativa entre os mapas de atenção e a interpretação de patologistas especialistas, o que limita a validação da explicabilidade do modelo. Além disso, ao observar os exemplos de mapas de atenção fornecidos, nota-se que em muitos dos casos o modelo se concentra em regiões que não são relevantes para a classificação de OED, como áreas de tecido conjuntivo e camada de queratina. Outro ponto a ser destacado é o elevado custo computacional e tempo de inferência dos modelos ViT, que podem ser desvantagens significativas em ambientes com recursos limitados.

Por fim, em Shephard et al. (2025), os autores propõem uma abordagem que utiliza *Vision Transformers* para a detecção de OED e predição de risco de malignidade. O modelo apresenta um F1-*score* de 0,96 na classificação entre tecidos displásicos e não displásicos e uma AUC de 0,73 na predição do risco de malignidade (baixo ou alto grau). Os autores também desenvolveram uma ferramenta computacional que realiza os dois níveis de classificação, além de permitir a visualização de regiões de atenção nas imagens histológicas. No entanto, assim como nos trabalhos anteriores, os autores não realizam uma análise comparativa entre os mapas de atenção gerados pelo modelo e a interpretação de patologistas. Outro ponto já mencionado são as desvantagens em termos de recursos computacionais associados aos modelos ViT.

1.2 OBJETIVOS

O objetivo geral deste trabalho é desenvolver uma plataforma educacional impulsionada por aprendizado profundo explicável voltada para o auxílio à formação de estudantes de pós-graduação em patologia bucal. A plataforma tem como objetivo a detecção de displasia epitelial oral a partir de imagens histológicas e dados clínicos de pacientes diagnosticados com leucoplasia.

1.2.1 Objetivos específicos

Os objetivos específicos deste trabalho são:

- Desenvolvimento de um sistema de aprendizado profundo para detecção de displasia epitelial oral a partir de imagens histológicas de casos de leucoplasia e dados complementares dos pacientes;
- Utilização de técnica visual de explicabilidade do processo de decisão do modelo de aprendizado profundo;

- Comparação dos mapas de explicabilidade gerados pelo modelo com a interpretação de patologistas especialistas;
- Desenvolvimento de uma ferramenta educacional *online* que utiliza o sistema proposto neste trabalho.

1.3 ORGANIZAÇÃO DO TRABALHO

O Capítulo 2 apresenta uma breve descrição a respeito de conceitos de patologia bucal, processamento de imagens, aprendizado de máquina e aprendizado profundo. No Capítulo 3, são detalhados os materiais e métodos utilizados no desenvolvimento deste trabalho. O Capítulo 4 apresenta os resultados obtidos juntamente com a discussão. Por fim, o Capítulo 5 apresenta a conclusão do trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, serão apresentados conceitos essenciais para compreensão do tema abordado neste trabalho.

2.1 DISPLASIA EPITELIAL ORAL

A leucoplasia oral, ilustrada na Figura 1, é definida como uma lesão esbranquiçada, potencialmente maligna e que não pode ser caracterizada clinicamente ou patologicamente como qualquer outra condição conhecida (Waal et al., 1997). Trata-se, portanto, de um diagnóstico exclusivamente clínico, baseado na exclusão de outras causas de placas brancas na mucosa oral. O procedimento padrão para a avaliação da leucoplasia consiste na realização de biópsia da região afetada, seguida de análise histopatológica. Durante a análise histopatológica, o tecido retirado da mucosa oral é processado e corado para permitir a visualização detalhada de suas camadas. A mucosa oral é composta principalmente por tecido epitelial e tecido conjuntivo. Em algumas regiões, o epitélio é coberto por uma camada de queratina. A Figura 2 fornece um exemplo de uma fotografia de uma lâmina histológica. Com base na análise histopatológica da amostra, a lesão pode ser classificada como com ou sem displasia epitelial.

A OED é uma lesão potencialmente maligna caracterizada por um conjunto de alterações arquiteturais e citológicas do tecido epitelial. Na classificação mais recente de tumores de cabeça e pescoço da OMS, há 28 características histológicas listadas, de maneira que o diagnóstico de displasia epitelial é feito pela identificação dessas características por um patologista (Muller; Tilakaratne, 2022). De acordo com Hankinson et al. (2024), alguns casos de OED são simples de diagnosticar, com anormalidades citológicas bem evidentes. No entanto, os autores argumentam que algumas displasias, aquelas que apresentam alterações arquiteturais como características predominantes, tornam-se casos mais desafiadores.

A gravidade da OED pode ser dividida, de acordo com a OMS (Muller; Tilakaratne, 2022), com base na extensão das alterações no epitélio: (i) leve, para alterações até o terço inferior; (ii) moderada, para alterações até o segundo terço; e (iii) severa, para alterações acima de dois terços do tecido epitelial. A Figura 3 ilustra a evolução da gravidade da displasia epitelial com base na presença e localização das células do epitélio (queratinócitos) atípicas. Entretanto, este procedimento de detecção de displasia e classificação em níveis (leve, moderada e severa) é altamente subjetivo, apresentando elevada variabilidade entre patologistas (Rocha et al., 2022; Alajaji et al., 2024; Peng et al., 2024).

2.2 PROCESSAMENTO DE IMAGENS DIGITAIS

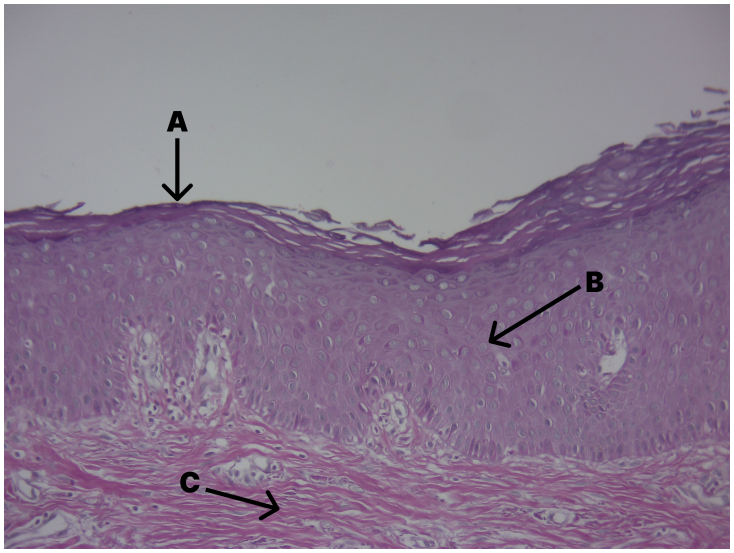
De acordo com Gonzalez e Woods (2018), uma imagem pode ser definida como uma função bidimensional $f(x,y)$, na qual x e y representam as coordenadas espaciais, e a amplitude de f em cada ponto corresponde à intensidade luminosa (ou nível de cinza) da imagem nesse

Figura 1 – Exemplo de leucoplasia.



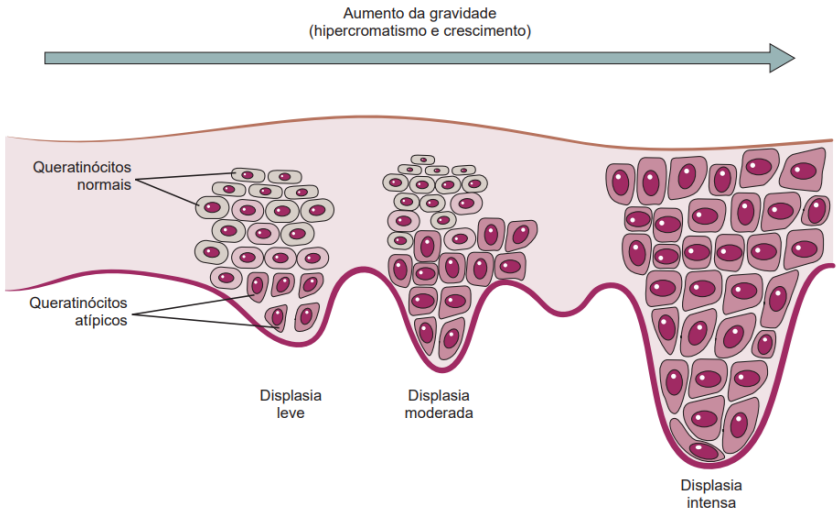
Fonte: Acervo interno - Laboratório de Patologia Bucal, Universidade Federal de Santa Catarina (2025).

Figura 2 – Exemplo de uma fotografia com ampliação de 200 vezes de uma lâmina histológica corada com hematoxilina e eosina referente a uma lesão bucal; em que “A” indica a camada de queratina, “B” o tecido epitelial e “C” o tecido conjuntivo.



Fonte: Acervo interno - Laboratório de Patologia Bucal, Universidade Federal de Santa Catarina (2025).

Figura 3 – Evolução da displasia epitelial.



Fonte: (Regezi; Sciubba; Jordan, 2017).

ponto. Quando x, y e os valores assumidos por f são finitos e discretos, a imagem é denominada imagem digital. Os elementos que compõem essa imagem são chamados de *pixels*, os quais representam a menor unidade de informação espacial.

As imagens digitais podem ser classificadas de acordo com a forma como os níveis de intensidade são representados: (i) imagens em níveis de cinza, compostas por um único canal; (ii) imagens coloridas, formadas por múltiplos canais, como por exemplo no modelo RGB; e (iii) imagens binárias, que contêm apenas dois valores possíveis (geralmente 0 e 1) e são frequentemente utilizadas para representar máscaras de regiões de interesse (ROI — do inglês, *Region of Interest*) ou resultados de segmentações.

O processamento de imagens digitais constitui um campo que engloba um conjunto de técnicas computacionais voltadas à manipulação, análise e interpretação de imagens para finalidades específicas, como melhoramento visual, extração de características, segmentação, reconhecimento de padrões ou suporte à tomada de decisão.

2.2.1 Morfologia matemática

A morfologia matemática é um conjunto de técnicas de processamento de imagens baseadas na teoria dos conjuntos, amplamente utilizada para a análise e manipulação da forma e estrutura de objetos presentes em imagens, principalmente binárias. Em vez de operar diretamente sobre valores de intensidade, as operações morfológicas atuam sobre a disposição espacial dos *pixels*, possibilitando transformações que preservam ou realçam propriedades geométricas dos objetos.

Essas operações utilizam um elemento estruturante — um conjunto de forma e tamanho definidos (por exemplo, um quadrado) — que percorre a imagem, interagindo com os *pixels* para modificar sua estrutura. O elemento estruturante determina como a operação morfológica

afetará os objetos. As operações morfológicas mais comuns são erosão e dilatação. Neste trabalho, será dado ênfase apenas à operação de dilatação, a qual é utilizada para expandir os objetos presentes em uma imagem binária.

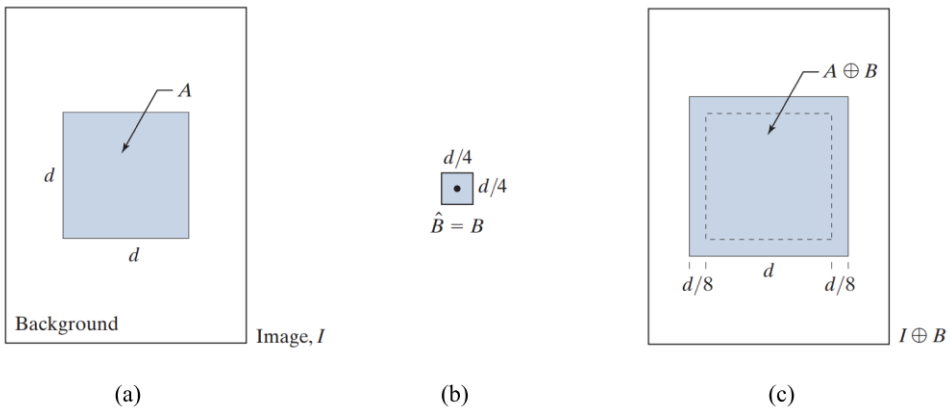
2.2.1.1 Dilatação

A dilatação é utilizada para expandir as regiões brancas (de valor 1) de uma imagem binária, fazendo com que objetos cresçam em tamanho e regiões próximas possam se unir. Seja A um conjunto que representa os pixels de valor 1 de uma imagem binária e B o elemento estruturante. A dilatação de A por B, denotada por $A \oplus B$, é definida como:

$$A \oplus B = \{z \mid (\hat{B})_z \cap A \neq \varnothing\} \tag{2.1}$$

em que $\hat{B}$ representa a reflexão de B em relação à sua origem e $(\hat{B})_z$ é o conjunto obtido ao transladar $\hat{B}$ para a posição z. Essa definição significa que um *pixel* pertencerá ao conjunto dilatado se a sobreposição da reflexão do elemento estruturante com a imagem original, naquela posição, tiver pelo menos um *pixel* em comum. A Figura 4 ilustra o efeito da dilatação.

Figura 4 – Exemplo de dilatação, em que (a) representa uma imagem I composta de um objeto A e seu fundo (*background*); (b) é o elemento estruturante B; e (c) é o resultado da dilatação de A por B, denotado por $A \oplus B$.



Fonte: Adaptado de Gonzalez e Woods (2018).

2.3 APRENDIZADO DE MÁQUINA E APRENDIZADO PROFUNDO

O campo da IA tem como objetivo o desenvolvimento de sistemas computacionais capazes de simular comportamentos inteligentes; como percepção e tomada de decisão. Nas últimas décadas, a IA vem evoluindo de abordagens baseadas em regras explícitas para métodos mais flexíveis e adaptativos, capazes de aprender diretamente a partir de dados. Essa evolução tem sido impulsionada pelo significativo aumento da disponibilidade de dados digitais, pela maior capacidade computacional e pelos avanços em algoritmos de processamento.

Nesse contexto, destacam-se o aprendizado de máquina e o aprendizado profundo, subáreas da IA voltadas para a criação de sistemas capazes de aprender padrões a partir de dados e tomar decisões ou realizar previsões.

O aprendizado de máquina é caracterizado por um conjunto de métodos que utilizam técnicas estatísticas para permitir que computadores aprendam com dados. O aprendizado ocorre por meio da identificação de padrões em conjuntos de exemplos, os quais são usados para construir modelos capazes de generalizar o conhecimento adquirido e aplicá-lo a dados ainda não vistos. Um dos principais tipos de aprendizado de máquina é o aprendizado supervisionado, onde o modelo é treinado com dados rotulados, ou seja, exemplos que possuem a resposta correta associada às entradas. O objetivo é que o modelo aprenda a mapear entradas para os rótulos corretos, permitindo previsões em novos dados.

Com o surgimento de grandes volumes de dados e o aumento da capacidade de processamento, tornou-se viável a aplicação de modelos mais complexos, dando origem ao chamado aprendizado profundo; um subconjunto do aprendizado de máquina. Um exemplo, são redes neurais artificiais compostas por múltiplas camadas. Esses modelos são especialmente eficazes em tarefas como reconhecimento de imagens, processamento de linguagem natural, detecção de padrões em sinais médicos, entre outras. Enquanto métodos clássicos de aprendizado de máquina frequentemente exigem um processo manual de extração de atributos, os modelos de aprendizado profundo têm a capacidade de aprender automaticamente representações diretamente dos dados brutos. Essa característica reduz a necessidade de intervenção humana na etapa de pré-processamento e pode aumentar o desempenho em tarefas complexas. Por outro lado, modelos de aprendizado profundo são conhecidos como "caixas-pretas", pois sua estrutura complexa dificulta a interpretação dos resultados e a compreensão dos padrões aprendidos. De maneira a mitigar essa limitação, técnicas de explicabilidade têm sido desenvolvidas para fornecer maior entendimento sobre como esses modelos tomam decisões (Chaddad et al., 2023).

Nos tópicos seguintes, serão abordados dois modelos de aprendizado de máquina: a máquina de vetores de suporte, um método clássico e amplamente utilizado de aprendizado de máquina raso; e as redes convolucionais, uma arquitetura de rede neural profunda amplamente utilizada em classificação de imagens.

2.3.1 Máquina de vetores de suporte

A máquina de vetores de suporte (SVM — do inglês, *Support Vector Machine*) é um método de aprendizado de máquina supervisionado amplamente utilizado em tarefas de classificação. Considerando um problema de classificação binária, o objetivo é encontrar um hiperplano ótimo que separe os dados das duas classes com a maior margem possível. Nesse sentido, a margem é definida como a distância entre o hiperplano de separação e os exemplos mais próximos de cada classe, chamados de vetores de suporte.

Para o caso de conjuntos de dados linearmente separáveis, o algoritmo busca resolver o seguinte problema de otimização:

$$\min_{\mathbf{w},b} \quad \frac{1}{2}\|\mathbf{w}\|^2 \tag{2.2}$$

$$\text{sujeito a} \quad y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \quad \forall i \tag{2.3}$$

em que $\mathbf{w}$ é o vetor de pesos que define o hiperplano, b é o termo de *bias* (viés), $y_i \in \{-1, 1\}$ é o rótulo da classe i e $\mathbf{x}_i$ é o vetor de atributos referente a i. A solução desse problema resulta no hiperplano que maximiza a margem entre as duas classes.

Entretanto, na grande maioria dos casos, os dados não são linearmente separáveis. Para lidar com essa situação, são introduzidas as chamadas variáveis de folga ξ_i, responsáveis por relaxar as imposições impostas ao problema de otimização. Assim, tem-se o SVM de margens suaves; cujo objetivo de otimização é então reformulado como:

$$\min_{\mathbf{w},b,\xi} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N} \xi_i \tag{2.4}$$

$$\text{sujeito a} \quad y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i \tag{2.5}$$

em que parâmetro $C > 0$ é um hiperparâmetro que controla o compromisso entre a margem máxima e o número de erros de classificação permitidos. Um valor alto de C penaliza fortemente os erros, tentando evitar violações, enquanto um valor menor permite maior flexibilidade e pode generalizar melhor em cenários com dados ruidosos.

A Figura 5 ilustra o conceito de SVM de margens suaves, mostrando como o hiperplano é ajustado para maximizar a margem entre as classes, mesmo quando os dados não são perfeitamente separáveis. Os pontos azuis representam uma classe do problema, enquanto que os triângulos vermelhos a outra classe. Os pontos e triângulos sem cores representam dados classificados erroneamente.

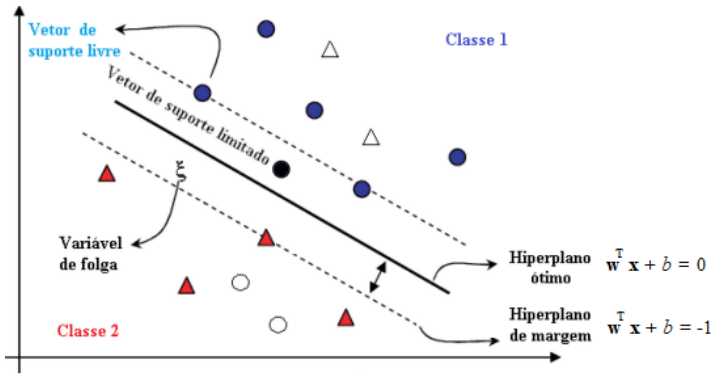
Além disso, o SVM pode ser estendido para problemas não linearmente separáveis por meio do uso de funções *kernel*, que transformam os dados de entrada em um espaço de atributos, onde se torna possível encontrar um hiperplano separador linear. Dentre os *kernels* mais utilizados estão o polinomial, o sigmoide e o *Radial Basis Function* (RBF).

2.3.2 Redes neurais artificiais

Uma rede neural artificial é composta por neurônios artificiais, também chamados de unidades, organizados em camadas. Cada neurônio recebe como entrada um vetor de atributos $\mathbf{x} = [x_1, x_2, \ldots, x_N]$, aplica uma transformação linear seguida de uma função de ativação. O modelo matemático de um neurônio é dado por:

$$z = \sum_{i=1}^{N} w_i x_i + b = \mathbf{w}^\top \mathbf{x} + b \tag{2.6}$$

Figura 5 – Representação do SVM de margens suaves.



Fonte: (Souza, 2022)

$$a = g(z) \tag{2.7}$$

em que $\mathbf{w}$ representa o vetor de pesos, b é o termo de *bias*, $g(\cdot)$ é a função de ativação e a é a saída do neurônio. A função de ativação introduz não-linearidades no modelo, permitindo que a rede aprenda relações complexas entre as entradas e saídas.

As redes mais simples são compostas por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída, organizadas em camadas totalmente conectadas (FC - do inglês, *fully connected*). Nesse tipo de arquitetura, cada neurônio de uma camada está conectado a todos os neurônios da camada seguinte, permitindo que a informação flua completamente entre as camadas. A camada de entrada recebe os dados brutos, enquanto as camadas ocultas realizam transformações progressivas nas representações internas, onde cada neurônio combina linearmente todas as saídas da camada anterior seguida de uma função de ativação não-linear, aprendendo assim padrões nos dados. A camada de saída fornece a predição final do modelo.

À medida que crescem a complexidade dos problemas e a disponibilidade de dados, torna-se necessário adotar modelos capazes de representar relações mais abstratas e sofisticadas. Nesse contexto, surgem as redes neurais profundas, caracterizadas pelo uso de múltiplas camadas ocultas entre a entrada e a saída. A adição de mais camadas à arquitetura permite que a rede aprenda representações hierárquicas dos dados, em que cada camada sucessiva captura características de nível mais alto; partindo de bordas e texturas em camadas iniciais, até formas e objetos mais complexos em camadas mais profundas (Bishop; Bishop, 2024). Essa profundidade permite que as redes neurais profundas sejam capazes de resolver tarefas complexas, como reconhecimento de imagens.

O treinamento de uma rede neural envolve o ajuste iterativo de seus pesos e vieses com o objetivo de minimizar uma função de perda (ou função custo), a qual mensura numericamente a discrepância entre as predições do modelo e os rótulos verdadeiros dos dados de treinamento. Em tarefas de classificação, uma das funções de perda mais utilizadas é a entropia cruzada, definida na Equação 2.8 (caso binário) e Equação 2.9 (caso multiclasse) (Géron, 2019).

$$\mathscr{L}_{BCE} = -\left[y \log(\hat{y}) + (1-y)\log(1-\hat{y})\right] \tag{2.8}$$

$$\mathscr{L}_{CCE} = -\sum_{k=1}^{K} y_k \log(\hat{y}_k) \tag{2.9}$$

em que y é o rótulo verdadeiro, $\hat{y}$ é a predição do modelo e K é o número de classes.

O processo de treinamento é organizado em épocas (*epochs*), onde cada época representa uma passagem completa por todo o conjunto de dados de treinamento. Durante cada época, os dados são divididos em pequenos lotes (*mini-batches*) de tamanho definido pelo hiperparâmetro tamanho do lote (*batch size*), que determina quantas amostras são processadas simultaneamente antes de atualizar os parâmetros da rede (Goodfellow; Bengio; Courville, 2016).

Em cenários onde a quantidade de dados disponível é limitada ou quando se busca acelerar o processo de treinamento, é comum empregar aprendizado por transferência (*transfer learning*) (Bishop; Bishop, 2024). Essa abordagem consiste em utilizar uma rede previamente treinada em uma tarefa relacionada — como a classificação de imagens em grandes conjuntos de dados — e reutilizar seus pesos como ponto de partida para uma nova tarefa. Frequentemente, congela-se parte das camadas iniciais da rede e realiza-se o ajuste fino (*fine-tuning*) das camadas finais, adaptando-as ao novo domínio. Isso permite reduzir o custo computacional, acelerar a convergência e melhorar o desempenho, especialmente quando o conjunto de dados da tarefa de interesse é pequeno.

Para encontrar os valores ótimos dos parâmetros da rede, empregam-se tipicamente métodos baseados na descida do gradiente, que atualizam os pesos na direção oposta ao gradiente da função de perda em relação a cada parâmetro, buscando reduzir o erro a cada iteração, isto é, ao processamento de um *mini-batch*. O cálculo eficiente desses gradientes é realizado por meio do algoritmo de retropropagação do erro (*backpropagation*) (Goodfellow; Bengio; Courville, 2016), que aplica sucessivamente a regra da cadeia do cálculo diferencial, propagando os gradientes do erro da camada de saída em direção às camadas anteriores da rede. Esse processo permite que os gradientes calculados em cada camada sejam usados para atualizar seus respectivos parâmetros de forma apropriada, fazendo com que a rede aprenda progressivamente representações mais eficazes para a tarefa em questão ao longo das múltiplas épocas de treinamento.

Apesar de sua eficácia, o treinamento de redes neurais profundas apresenta diversos desafios práticos. Um dos problemas mais significativos é o desaparecimento do gradiente, que ocorre quando os gradientes se tornam exponencialmente menores conforme são propagados através das camadas mais profundas da rede durante a retropropagação. Esse fenômeno resulta em atualizações muito pequenas nos pesos das camadas iniciais, impedindo o aprendizado efetivo de características importantes nos estágios iniciais do processamento. Para mitigar esses problemas, diversas técnicas de regularização e arquiteturas especializadas foram desenvolvidas (Bishop; Bishop, 2024).

A normalização em lote (*batch normalization*) (Ioffe; Szegedy, 2015) é uma técnica fundamental que estabiliza o treinamento ao padronizar as entradas de cada camada, ajustando os valores para que tenham média zero e variância unitária dentro de cada *mini-batch*. Isso resolve o problema da mudança constante na distribuição dos dados internos da rede, onde os valores das ativações variam continuamente durante o treinamento, tornando o processo mais lento e instável (Ioffe; Szegedy, 2015). Além disso, arquiteturas como as redes residuais (ResNets) introduzem conexões que permitem que o gradiente flua diretamente através das camadas, facilitando o treinamento de redes muito profundas (He et al., 2015).

2.3.3 Redes convolucionais

As redes neurais convolucionais (CNNs — do inglês, *Convolutional Neural Networks*) foram desenvolvidas especificamente para processar dados que possuem estrutura espacial, sendo particularmente eficazes no processamento de imagens. A motivação para seu desenvolvimento surge das limitações das redes neurais totalmente conectadas quando aplicadas a tarefas de visão computacional. Ao contrário das redes neurais totalmente conectadas, as CNNs exploram a organização espacial dos dados por meio de operações que consideram a vizinhança local dos elementos, reduzindo significativamente o número de parâmetros e promovendo o aprendizado de características mais relevantes e generalizáveis. Em se tratando de imagens, a proximidade espacial entre *pixels* carrega informação semântica importante: *pixels* adjacentes tendem a ter valores correlacionados e formar padrões locais como bordas, texturas e formas. As CNNs foram projetadas para explorar essas propriedades espaciais de forma eficiente, preservando as relações de vizinhança e detectando padrões locais que podem ocorrer em qualquer posição da imagem.

A arquitetura típica de uma CNN é composta por uma sequência de camadas convolucionais, camadas de ativação, camadas de amostragem (*pooling*), e eventualmente camadas totalmente conectadas nas etapas finais para classificação.

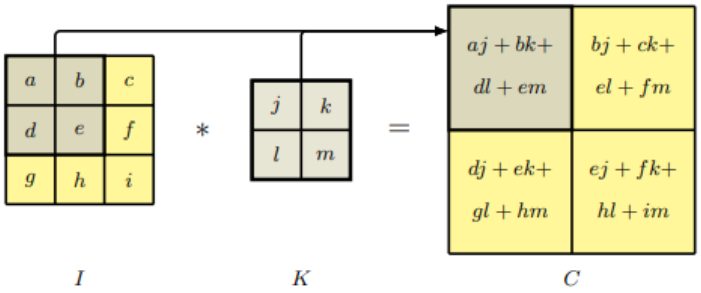
2.3.3.1 Camadas convolucionais e de ativação

Cada camada convolucional aplica múltiplos filtros à entrada, produzindo mapas de atributos (*feature maps*) que destacam diferentes aspectos da imagem de entrada. A convolução entre uma imagem $\mathbf{I}$ com valores de *pixels* $I(j,k)$ e um filtro $\mathbf{K}$ com valores de *pixels* $K(l,m)$ (Figura 6) é definida como:

$$C(j,k) = \mathbf{I} * \mathbf{K} = \sum_{l} \sum_{m} I(j+l, k+m) K(l,m) \tag{2.10}$$

Cada filtro convolucional é uma matriz pequena (tipicamente 3×3 ou 5×5) de pesos treináveis que detecta padrões específicos na imagem. Durante a operação de convolução, o filtro desliza sobre toda a extensão da imagem de entrada com um passo definido, computando

Figura 6 – Exemplo de uma imagem 3×3 convoluída com um filtro 2×2, resultando em um mapa 2×2.



Fonte: (Bishop; Bishop, 2024)

o produto escalar entre os valores do filtro e a região local da imagem sobreposta. Esse processo resulta em um mapa de atributos que indica a presença e intensidade do padrão detectado pelo filtro em cada posição da imagem.

Uma camada convolucional típica aplica múltiplos filtros independentes à entrada, produzindo múltiplos mapas de atributos. As camadas de ativação são aplicadas após a convolução para introduzir não-linearidades no modelo, permitindo que a rede aprenda relações mais complexas. A unidade retificada linear (ReLu — *Rectified Linear Unit*) é uma das funções de ativação mais comuns das camadas ocultas, definida como:

$$g(\mathbf{Z}) = \max(0, Z(i,j)) \tag{2.11}$$

2.3.3.2 Camadas de pooling

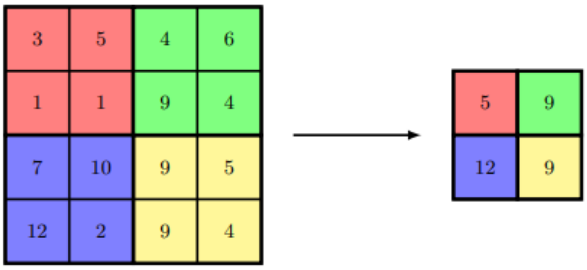
As camadas de *pooling* são utilizadas para reduzir a dimensionalidade dos mapas de atributos, mantendo as informações mais relevantes e diminuindo o custo computacional. O tipo mais comum é a *max pooling* (Figura 7), que seleciona o maior valor em uma região do mapa de atributos. Essa operação contribui para a invariância a pequenas transformações na entrada, como translações e ruídos locais, além de diminuir a propensão ao *overfitting*.

O treinamento de redes convolucionais segue o mesmo princípio das redes neurais tradicionais: busca-se minimizar uma função de perda, geralmente por meio do algoritmo de descida do gradiente ou suas variantes. O algoritmo *backpropagation* é adaptado para considerar as operações de convolução e *pooling*, calculando os gradientes dos filtros de forma eficiente. Durante o treinamento, os filtros são atualizados iterativamente.

2.3.3.3 Camadas de saída

As camadas de saída são responsáveis por produzir a saída final da rede neural. Em tarefas de classificação, a camada de saída geralmente consiste em uma camada totalmente conectada seguida por uma função de ativação, como a sigmoide (Equação 2.12) ou a *softmax*

Figura 7 – Ilustração da operação de *max pooling*, no qual blocos de 2×2 *pixels* em um mapa de atributos são combinados usando o operador *max* para gerar um novo mapa de atributos de menor dimensionalidade.



Fonte: (Bishop; Bishop, 2024)

(Equação 2.13); sendo a primeira utilizada para problemas de classificação binária e a segunda para problemas de classificação multiclasse.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{2.12}$$

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{2.13}$$

2.3.4 ResNet

À medida que redes neurais profundas passaram a ser utilizadas para tarefas mais complexas, observou-se que simplesmente aumentar a profundidade da rede nem sempre resultava em melhor desempenho. Na prática, redes muito profundas apresentavam dificuldades de treinamento, como o desaparecimento do gradiente. De forma a enfrentar esse problema, He et al. (2015) propuseram a arquitetura de redes residuais (ResNets), que introduz conexões residuais (atalhos), que permitem que o sinal de entrada de uma camada seja somado diretamente à sua saída.

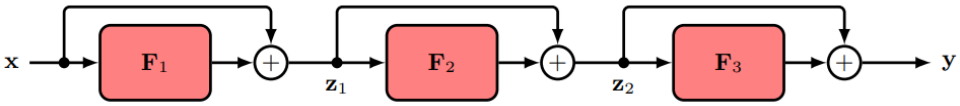
Considerando a Figura 8, um bloco residual pode ser representado por:

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \mathbf{x} \tag{2.14}$$

em que $\mathbf{x}$ é a entrada do bloco, $\mathbf{F}(\mathbf{x})$ representa a transformação aprendida pelas camadas dentro do bloco (geralmente uma sequência de camadas convolucionais, *batch normalization* e funções de ativação), e $\mathbf{y}$ é a saída do bloco. A adição da entrada $\mathbf{x}$ à saída da transformação $\mathbf{F}(\mathbf{x})$ cria um atalho que facilita o fluxo do gradiente durante o treinamento. O termo "residual" vem do fato de que, em cada bloco, a função $\mathbf{F}$ aprende apenas a diferença residual entre a entrada e a saída desejada (Bishop; Bishop, 2024).

Essa arquitetura permitiu o treinamento de redes muito mais profundas. As ResNets demonstraram desempenho superior às arquiteturas anteriores em diversas tarefas de visão computacional (He et al., 2015).

Figura 8 – Exemplo de uma rede residual com três blocos residuais.



Fonte: (Bishop; Bishop, 2024).

2.4 MECANISMOS DE ATENÇÃO E INTERPRETABILIDADE EM REDES CONVOLUCIONAIS

Nos tópicos a seguir, serão apresentados os principais conceitos relacionados aos métodos de localização de regiões relevantes e interpretabilidade em redes convolucionais, abordando inicialmente os métodos baseados em mapas de ativação e, em seguida, as arquiteturas com mecanismos de atenção explícitos. Esses conceitos fornecem a base teórica necessária para a compreensão do modelo proposto neste trabalho e das estratégias adotadas para a análise e interpretação de seus resultados.

2.4.1 Método CAM

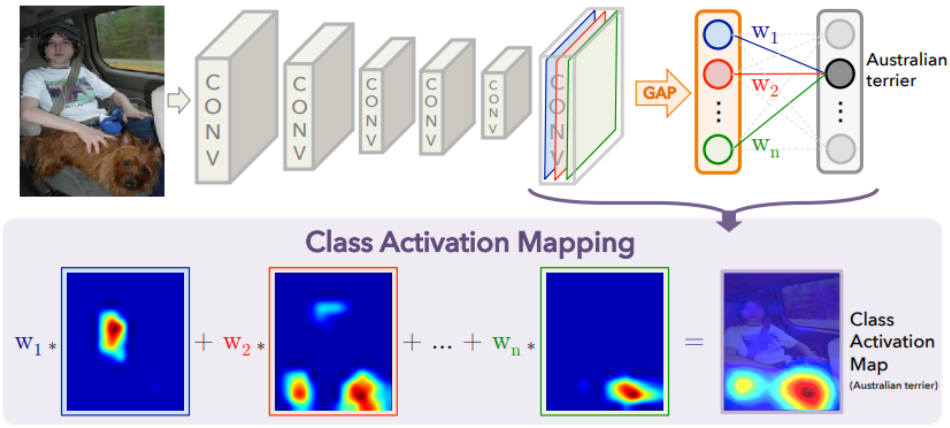
A utilização de modelos de aprendizado profundo também trouxe a necessidade de desenvolver técnicas que permitam a interpretação e explicação das decisões tomadas por esses modelos (Chaddad et al., 2023). Nesse contexto, Zhou et al. (2015) propuseram o método CAM (*Class Activation Mapping*), que permite visualizar as regiões da imagem mais relevantes para a predição de uma determinada classe. A técnica baseia-se na capacidade das camadas convolucionais de preservar informações espaciais. O método CAM explora essa propriedade para gerar mapas de ativação que destacam as áreas discriminativas para cada classe.

A arquitetura necessária para a aplicação do CAM consiste em uma rede convolucional seguida de uma camada de média global (GAP — do inglês, *Global Average Pooling*) e, por fim, uma camada totalmente conectada que realiza a classificação. A camada GAP calcula a média espacial de cada mapa de atributos da última camada convolucional, reduzindo cada mapa a um único valor. Esses valores são então ponderados pelos pesos da camada de classificação para produzir a pontuação (*score*) final de cada classe.

O mapa de ativação para uma classe específica c é obtido através de uma combinação linear dos mapas de atributos da última camada convolucional ponderados pelos pesos correspondentes à classe c na camada totalmente conectada, conforme ilustrado na Figura 9. A Equação 2.15 descreve esse processo, onde g_k representa o k-ésimo mapa de atributos, e w_k^c é o peso associado à classe c para o mapa g_k.

$$M_c(x,y) = \sum_k w_k^c g_k(x,y) \tag{2.15}$$

Figura 9 – Exemplo de implementação do método CAM.



Fonte: (Zhou et al., 2015).

No entanto, essa abordagem apresenta uma limitação importante: ela depende de uma estrutura arquitetural específica, o que restringe sua aplicabilidade a redes já treinadas com esse formato. Essa restrição motivou o desenvolvimento de métodos mais flexíveis, como o Grad-CAM, que será apresentado a seguir.

2.4.2 Método Grad-CAM

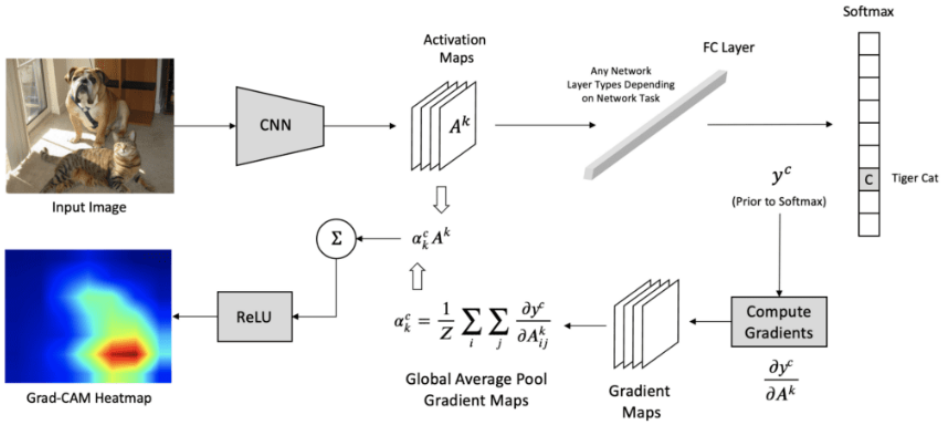
Com o objetivo de superar a limitação do método CAM e tornar a técnica de explicabilidade compatível com uma gama mais ampla de arquiteturas, o Grad-CAM (*Gradient-weighted Class Activation Mapping*) — proposto por Selvaraju et al. (2017) — destaca-se como uma técnica de explicabilidade de modelos de redes convolucionais. O método utiliza os gradientes da saída correspondente à classe prevista com relação às ativações da última camada convolucional da rede para gerar um mapa de calor (*heatmap*) referente à imagem de entrada. Esse mapa destaca visualmente as áreas que mais influenciaram a decisão do modelo.

Essa técnica calcula a média global (GAP) do gradiente de determinado *score* de classificação da classe de saída y^c em relação ao mapa de atributos A_k da última camada convolucional da rede, conforme Equação 2.16, em que Z é o número de elementos do mapa de atributos. Em seguida, é feita uma combinação ponderada com o mapa de atributos, seguida de uma função ReLU, de acordo com a Equação 2.17. A Figura 10 ilustra o processo de implementação do método. O resultado obtido é um mapa de calor que pode ser sobreposto à imagem original para facilitar a interpretação visual, destacando as áreas que mais contribuíram para a predição da classe.

$$\alpha_k^c = \frac{1}{Z}\sum_i\sum_j \frac{\partial y^c}{\partial A_{i,j}^k} \tag{2.16}$$

$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A_k\right) \tag{2.17}$$

Figura 10 – Exemplo de implementação do algoritmo Grad-CAM.



Fonte: (Dawn, 2023).

2.4.3 Attention Branch Networks

Embora abordagens como o Grad-CAM forneçam visualizações geradas após o modelo ser treinado, elas não influenciam diretamente o aprendizado do modelo durante o treinamento. Isso motivou o desenvolvimento de arquiteturas capazes de integrar mecanismos de atenção de forma explícita e conjunta ao processo de treinamento, resultando em modelos mais interpretáveis e, potencialmente, mais eficazes.

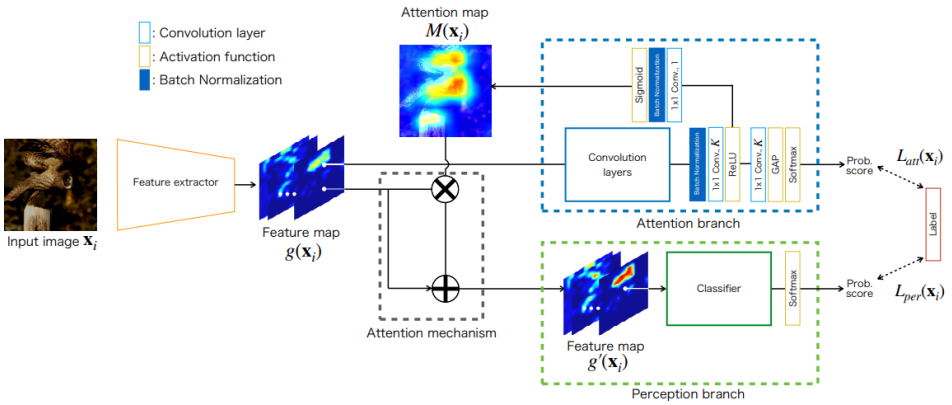
Nesse contexto, Fukui et al. (2018) propuseram a *Attention Branch Network* (ABN), uma arquitetura que une atenção visual e aprendizado supervisionado de maneira integrada (Figura 11). A ideia central é permitir que a rede destaque, de forma supervisionada, as regiões mais relevantes da imagem para a tarefa de classificação, ao mesmo tempo em que utiliza essas regiões para melhorar seu próprio desempenho preditivo. Assim, a ABN combina explicabilidade e desempenho em um único modelo, fornecendo mapas de atenção semelhantes ao Grad-CAM, mas de forma treinável.

A arquitetura da ABN é baseada em quatro componentes principais: extrator de atributos (também chamado de *feature extractor* ou *backbone*), o ramo de atenção (*attention branch*), o mecanismo de atenção (*attention mechanism*) e o ramo de percepção (*perception branch*).

2.4.3.1 Extrator de atributos

O extrator de atributos é responsável por extrair características relevantes da imagem de entrada. Normalmente é utilizada uma rede convolucional pré-treinada, como a ResNet, que converte a imagem em um mapa de atributos de alta dimensão. Para isso, são removidas as camadas finais de classificação da rede, permitindo que o extrator produza um mapa de características que captura informações visuais importantes. A saída do extrator é um mapa de atributos $g(\mathbf{x})$ de dimensões $H \times W \times C$, onde H e W são altura e largura espacial, e C é o número de canais.

Figura 11 – Exemplo de implementação da arquitetura ABN.



Fonte: (Fukui et al., 2018).

2.4.3.2 *Ramo de atenção*

O ramo de atenção é responsável por gerar um mapa de atenção que destaca as regiões mais relevantes da imagem para a tarefa de classificação. Ele recebe o mapa de atributos do extrator e gera dois elementos: (i) um mapa de atenção supervisionado, que realça as regiões mais relevantes da imagem para a tarefa de classificação e; (ii) uma predição auxiliar de classe, usada para guiar o aprendizado da atenção com base nas classes reais.

Para gerar o mapa de atenção $M(\mathbf{x})$ de dimensões $H \times W \times 1$, este ramo aplica uma sequência de camadas convolucionais seguidas de uma ativação sigmoide, produzindo uma máscara com valores entre 0 e 1. Essa máscara é treinada supervisionadamente de acordo com a função de perda definida em sequência.

2.4.3.3 *Mecanismo de atenção*

O mecanismo de atenção utiliza o mapa de atenção gerado pelo ramo de atenção para modificar o mapa de atributos extraído. A modificação é feita por meio de uma operação de multiplicação elemento a elemento entre o mapa de atributos $g(\mathbf{x})$ e o mapa de atenção $M(\mathbf{x})$, conforme a Equação 2.18:

$$g'(\mathbf{x}) = g(\mathbf{x}) \cdot (1 + M(\mathbf{x})) \tag{2.18}$$

em que $g'(\mathbf{x})$ é o mapa de atributos modificado, que enfatiza as regiões destacadas pelo mapa de atenção. O termo $1 + M(\mathbf{x})$ garante que as regiões não destacadas ainda contribuam para a predição, mas com menor peso.

2.4.3.4 Ramo de percepção

O ramo de percepção é responsável por realizar a classificação final da imagem. O classificador utilizado nesse ramo consiste de camadas totalmente conectadas que recebem o mapa de atributos modificado $g'(\mathbf{x})$ como entrada e produzem a predição final da classe.

2.4.3.5 Função de perda composta

A ABN utiliza uma função de perda composta que otimiza os dois ramos simultaneamente. A função de perda total é definida como a soma da perda do ramo de atenção e da perda do ramo de percepção, ponderadas por um fator λ, onde $\lambda \in [0,1]$, conforme a Equação 2.19:

$$\mathscr{L}_{\text{total}} = \lambda \mathscr{L}_{\text{attention}} + (1-\lambda)\mathscr{L}_{\text{perception}} \tag{2.19}$$

em que $\mathscr{L}_{\text{attention}}$ é definida através da entropia cruzada entre a predição auxiliar do ramo de atenção e as classes reais e $\mathscr{L}_{\text{perception}}$ é definida como a entropia cruzada entre a predição final do ramo de percepção e as classes reais.

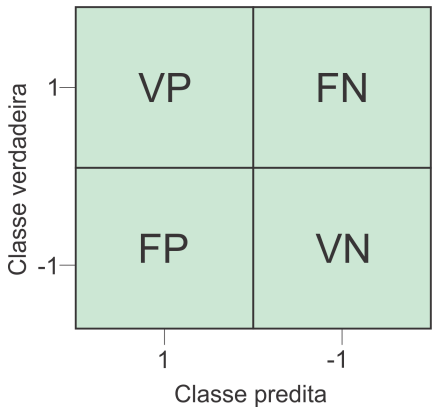
2.5 AVALIAÇÃO DE DESEMPENHO

A avaliação de desempenho é uma etapa fundamental no desenvolvimento de modelos de aprendizado de máquina, permitindo quantificar a qualidade das predições. Para modelos de classificação, esta avaliação é baseada na análise das predições corretas e incorretas realizadas pelo modelo em um conjunto de dados de teste; isto é, não vistos pelo modelo durante seu treinamento. A seguir serão detalhadas as principais métricas de avaliação de desempenho.

2.5.1 Matriz de confusão

A matriz de confusão é uma representação tabular que organiza a quantidade de amostras classificadas em cada categoria do problema. Em um cenário de classificação binária, com uma classe positiva e outra negativa, a matriz apresenta dimensão 2×2. Nesse caso, as linhas correspondem às classes reais, enquanto as colunas representam as classes previstas pelo modelo, permitindo uma análise clara do desempenho do classificador em termos de acertos e erros. A Figura 12 ilustra a matriz de confusão para um problema de classificação binária, em que VP (Verdadeiro Positivo) representa o número de amostras corretamente classificadas como positivas, VN (Verdadeiro Negativo) representa o número de amostras corretamente classificadas como negativas, FP (Falso Positivo) representa o número de amostras incorretamente classificadas como positivas e FN (Falso Negativo) representa o número de amostras incorretamente classificadas como negativas.

Figura 12 – Matriz de confusão para um problema de classificação binária.



Fonte: (Souza, 2022).

2.5.2 Métricas de desempenho

A partir da matriz de confusão, algumas métricas de desempenho podem ser obtidas, tais como:

$$\text{Acurácia} = \frac{VP+VN}{VP+VN+FP+FN} \tag{2.20}$$

$$\text{Precisão} = \frac{VP}{VP+FP} \tag{2.21}$$

$$\text{Sensibilidade} = \frac{VP}{VP+FN} \tag{2.22}$$

$$\text{Especificidade} = \frac{VN}{VN+FP} \tag{2.23}$$

$$F_1 = 2 \cdot \frac{\text{Precisão} \cdot \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}} \tag{2.24}$$

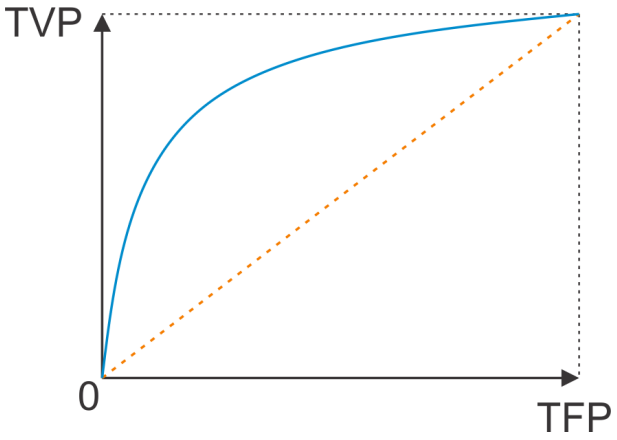
em que:

- Acurácia: proporção de previsões corretas em relação ao total de amostras;
- Precisão: proporção de verdadeiros positivos dentre todas as amostras classificadas como positivas;
- Sensibilidade: proporção de verdadeiros positivos detectados em relação ao total de casos positivos existentes. Em contextos clínicos, representa a capacidade de identificar corretamente os pacientes doentes;
- Especificidade: proporção de verdadeiros negativos detectados em relação ao total de casos negativos existentes. Em contextos clínicos, representa a capacidade de identificar corretamente os pacientes saudáveis;

- F1-*Score*: média harmônica entre precisão e sensibilidade. É útil quando há necessidade de equilíbrio entre identificar corretamente casos positivos e evitar falsos positivos.

Uma forma amplamente utilizada para avaliar o desempenho de classificadores binários é por meio da curva característica de operação do receptor (ROC — do inglês, *Receiver Operating Characteristic*). Essa curva é um gráfico (Figura 13) que relaciona a taxa de verdadeiros positivos (TVP) com a taxa de falsos positivos (TFP) ao variar os limiares de decisão do classificador. A AUC fornece uma medida quantitativa desse desempenho, assumindo valores entre 0 e 1, em que valores mais próximos de 1 indicam maior capacidade de discriminar corretamente entre as classes.

Figura 13 – Curva ROC (linha contínua azul) para um problema de classificação binária. Em laranja pontilhado é apresentado o resultado de um classificador aleatório.



Fonte: (Souza, 2022).

2.5.3 Métricas de avaliação dos mapas de atenção

Para avaliar a qualidade dos mapas de atenção gerados por modelos interpretáveis, como o Grad-CAM, podem ser utilizadas métricas que comparam esses mapas com anotações de referência feitas por especialistas. Métricas comumente utilizadas para avaliação de segmentação de imagens (Müller; Soto-rey; Kramer, 2022) não são as melhores escolhas para esse propósito, uma vez que os mapas de atenção não são voltados para segmentação de imagens, mas sim para representações que indicam a importância relativa das regiões da imagem.

O trabalho de Zhang et al. (2017) apresenta a métrica *Pointing Game*, que avalia a capacidade do mapa de atenção em destacar corretamente as regiões relevantes da imagem. A métrica é calculada da seguinte forma: primeiro, identifica-se o ponto de maior ativação no mapa de atenção. Em seguida, verifica-se se esse ponto está localizado dentro da região anotada como relevante pelos especialistas. Se estiver, conta-se como um acerto; caso contrário, conta-se como um erro. A acurácia do *Pointing Game* é então calculada como a razão entre o número de acertos e o número de imagens avaliadas.

Em Desai e Ramaswamy (2020), os autores propõem uma avaliação baseada na comparação dos K% *pixels* mais ativados no mapa de atenção. Essa abordagem considera apenas uma fração dos *pixels* mais relevantes. A partir dessa ideia, pode-se calcular a métrica denominada de Precisão Top-K%, que calcula a porcentagem de *pixels* entre os K% mais ativados no mapa de atenção que estão localizados dentro das regiões de interesse. Essa métrica oferece uma avaliação mais robusta da capacidade do mapa de atenção em destacar as áreas relevantes da imagem, considerando múltiplos pontos de ativação em vez de apenas o ponto máximo.

Além disso, o presente trabalho propõe a métrica Fração de Ativação, que mede a proporção da soma das ativações dentro das regiões anotadas pelos especialistas em relação à soma total das ativações no mapa de atenção. Essa métrica fornece uma visão quantitativa da concentração das ativações nas áreas de interesse, permitindo avaliar o quão bem o mapa de atenção foca nas regiões relevantes para a tarefa em questão.

3 METODOLOGIA

Este capítulo descreve os materiais e métodos utilizados neste trabalho, especificando a metodologia de detecção de OED e sua utilização na plataforma educacional proposta.

3.1 BANCO DE DADOS

O presente trabalho foi aprovado pelo Comitê de Ética em Pesquisa com Seres Humanos da UFSC (Plataforma Brasil, CAAE 15025319.3. 0000.0121). O banco de dados utilizado consiste em imagens histopatológicas e dados complementares dos pacientes do Biobanco do Laboratório de Patologia Bucal da UFSC (CONEP B-051; Processo Nº. 25000.237810/2014-54).

3.1.1 Banco de imagens

O banco de dados utilizado neste estudo é composto por 114 casos clínicos de pacientes diagnosticados com leucoplasia oral, totalizando 406 imagens histopatológicas com resolução de 2584×1936 *pixels*. As imagens foram obtidas a partir de uma ampliação de 200 vezes de lâminas coradas com hematoxilina e eosina, cujas seções representativas foram digitalizadas por meio de um microscópio biológico trinocular da marca Novel BM2000 com câmera para captura de imagens marca Tucsen ISH500 de 5.0 MP acoplada; constituindo as imagens do banco de dados. Entre os casos de leucoplasia oral, 59 apresentam displasia epitelial, correspondendo a 193 imagens, enquanto os 55 casos restantes não apresentam displasia, totalizando 213 imagens. Os casos que não apresentam displasia foram diagnosticados como hiperceratose e acantose (Bruch; Treister, 2010). Como o objetivo do trabalho é a detecção de displasia, as classes utilizadas foram: displásico e não displásico.

A limitação de dados constitui um desafio recorrente no contexto da pesquisa em saúde (Rocha et al., 2025). Ademais, o processo de anotação de imagens histopatológicas demanda a atuação de patologistas qualificados, recurso frequentemente limitado em ambientes de pesquisa. Esses fatores combinados dificultam a construção de conjuntos de dados extensos e diversificados. Nesse contexto, o banco de dados empregado neste estudo, embora apresente limitações quantitativas, representa uma contribuição relevante para o avanço da pesquisa na área. Tal afirmação fundamenta-se no fato de que o conjunto de imagens e casos compilado supera em dimensão outros bancos de dados utilizados em investigações recentes similares (Rocha et al., 2025), os quais já se destacavam por ampliar substancialmente o volume de imagens em relação a trabalhos anteriores.

3.1.2 Dados complementares dos pacientes

O estudo apresentado em Rocha et al. (2025) demonstrou que a incorporação de dados complementares dos pacientes pode aumentar significativamente o desempenho de modelos de aprendizado de máquina na detecção de displasia epitelial. Com base nesse resultado, o banco de dados utilizado neste trabalho inclui, além das imagens histopatológicas, informações clínicas complementares dos pacientes, a saber: característica clínica da lesão, localização anatômica da lesão, idade, histórico de tabagismo e de etilismo. Esses atributos foram selecionados por dois patologistas orais com base em sua relevância diagnóstica, sendo apontados como os fatores com maior influência na identificação de displasia epitelial (Rocha et al., 2025). Para viabilizar a integração desses dados ao modelo, os atributos foram convertidos em variáveis binárias, indicando a presença ou ausência de fatores de risco associados. A transformação foi realizada conforme a metodologia proposta por Rocha et al. (2025), descrita a seguir:

- Característica clínica da lesão: valor "1" atribuído para lesões heterogêneas; "0" para as demais;
- Localização da lesão: valor "1" para lesões localizadas na borda lateral da língua, ventre da língua ou assoalho bucal; "0" para outras localizações;
- Idade do paciente: valor "1" para pacientes com idade superior a 40 anos; "0" caso contrário;
- Histórico de tabagismo: valor "1" para pacientes fumantes ou ex-fumantes; "0" para não fumantes;
- Histórico de etilismo: valor "1" para pacientes etilistas ou ex-etilistas; "0" para não etilistas.

3.1.3 Redimensionamento de imagens e separação de conjuntos de dados

Inicialmente, as imagens foram redimensionadas devido a limitações computacionais. A nova resolução adotada foi de 700×524 *pixels*, mantendo-se a mesma razão da dimensão original.

Em sequência, o banco de dados foi particionado em conjuntos de treinamento e teste, assegurando que imagens pertencentes a um mesmo paciente fossem alocadas exclusivamente a um único conjunto, evitando vazamento de informação. Foi utilizada uma separação de 70% para o conjunto de treinamento e 30% para o conjunto de teste. Sendo assim, o conjunto de treinamento apresenta 79 casos; sendo 39 com displasia (totalizando 132 imagens) e 40 sem displasia (totalizando 156 imagens). O conjunto de teste apresenta 35 casos; sendo 20 com displasia (totalizando 61 imagens) e 15 sem displasia (totalizando 57 imagens).

3.2 ANOTAÇÃO DAS REGIÕES DE INTERESSE E GERAÇÃO DOS MAPAS DE ATENÇÃO DE REFERÊNCIA

Um dos objetivos do trabalho foi comparar os mapas de atenção gerados pelo modelo com a percepção de um patologista das regiões de interesse nas imagens histopatológicas. Para isso, anotações manuais das regiões de interesse foram realizadas por um patologista oral e revisadas por um segundo patologista oral, servindo como referência para a avaliação dos mapas de atenção gerados pelo modelo. O critério adotado para a anotação foi a identificação de regiões com características histológicas sugestivas de displasia epitelial. Para casos em que não há displasia, foi demarcada a região completa correspondente ao tecido epitelial.

Devido à natureza exaustiva do processo de anotação manual das regiões de interesse, optou-se por anotar uma amostra representativa do conjunto de dados. Para o conjunto de treinamento, foram anotadas 89 imagens (aproximadamente 30% do total), sendo 45 não displásicas e 44 displásicas. A seleção dessas imagens priorizou dois aspectos: (i) a viabilidade operacional do processo de anotação, que demanda análise minuciosa por patologista especialista, e (ii) a garantia de representatividade equilibrada dos diferentes graus de displasia (leve, moderada e severa), assegurando diversidade nos padrões histopatológicos utilizados para supervisão espacial. Para o conjunto de teste, todas as imagens foram anotadas, permitindo uma avaliação abrangente da qualidade dos mapas de atenção gerados pelo modelo.

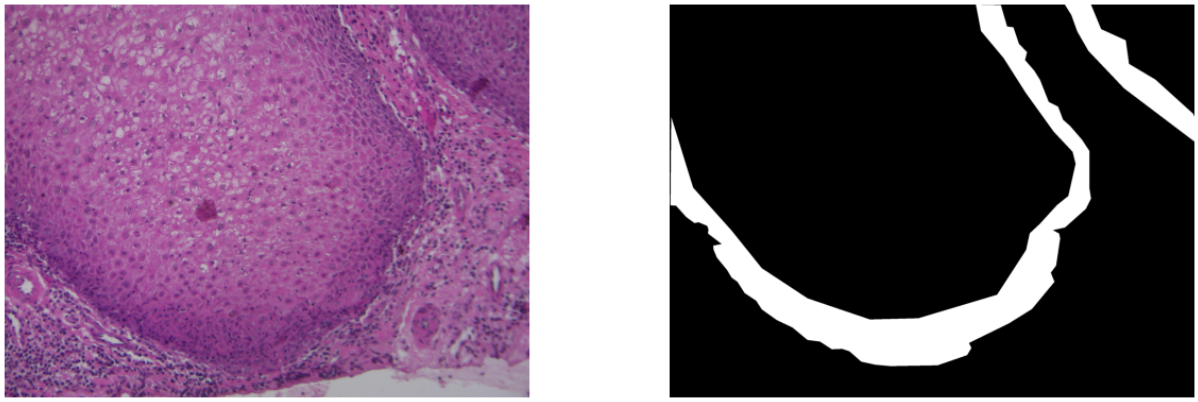
O procedimento de anotação se deu através do *software Roboflow* (Dwyer et al., 2025). Para as imagens de displasia, foram consideradas duas regiões: região displásica e restante da imagem (tecido epitelial não displásico, tecido conjuntivo e fundo da lâmina). Para imagens não displásicas, foi anotada a região completa correspondente ao tecido epitelial e o restante da imagem (tecido conjuntivo e fundo da lâmina). Como mencionado previamente, a região displásica sempre estará contida na região do tecido epitelial.

Após a demarcação das regiões, as anotações foram exportadas para um arquivo de texto contendo as coordenadas e classes das regiões anotadas. Os mapas de atenção foram gerados a partir dessas informações alocadas pelo patologista, de modo que regiões displásicas recebessem valor 1 e o restante das regiões recebesse valor 0. Dessa forma, os mapas de atenção foram gerados como máscaras binárias; onde regiões displásicas são representadas por valores mais altos (cor branca), enquanto que o restante das regiões é representado por valores baixos (cor preta). A Figura 14 apresenta um exemplo de imagem histopatológica e seu mapa de atenção gerado a partir das anotações do patologista.

3.3 METODOLOGIA DE DETECÇÃO DE DISPLASIA EPITELIAL ORAL

Este trabalho propõe um modelo para a detecção de displasia epitelial oral que integra informações em dois níveis: imagens histopatológicas e dados complementares dos pacientes, seguindo a metodologia de Rocha et al. (2025). A seção 3.3.1 descreve o que foi denominado de modelo base, que consiste em uma ResNet50 treinada para a detecção de displasia epitelial oral

Figura 14 – Exemplo de mapa de atenção: (a) Exemplo de imagem histopatológica; (b) Mapa de atenção da imagem histopatológica.



(a) (b)

Fonte: do autor.

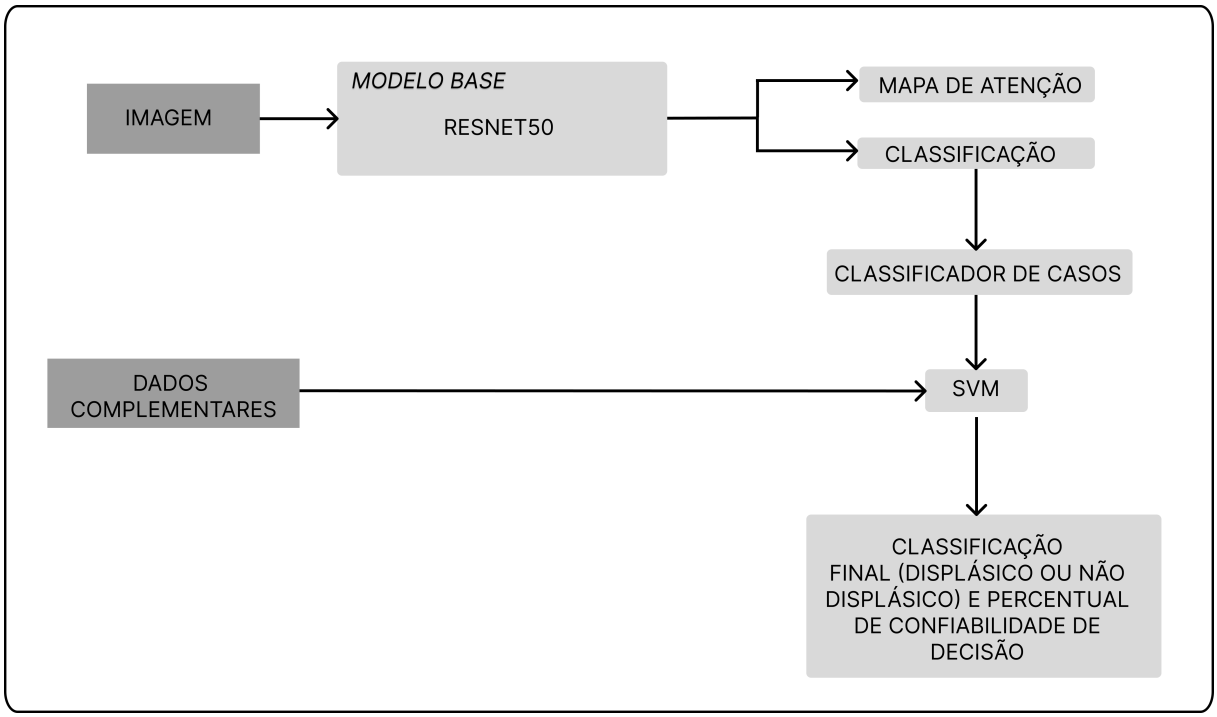
nas imagens histopatológicas. A seção 3.3.2 descreve o modelo ABN modificado, que utiliza uma ResNet50 como extrator de atributos e incorpora um mecanismo de atenção. A seção 3.3.3 descreve como as duas abordagens anteriores são complementadas por um classificador SVM, que integra a pontuação de classificação obtida pelo modelo de aprendizado profundo com os dados complementares dos pacientes para fornecer a classificação final.

3.3.1 Modelo base

A Figura 15 apresenta o diagrama em blocos da metodologia proposta para detecção de displasia epitelial oral utilizando o modelo base.

Primeiramente, uma ResNet50 é utilizada para realizar a detecção de displasia a partir de uma dada imagem. O treinamento desta rede consistiu na utilização de aprendizado por transferência, em que se utilizou uma ResNet50 pré-treinada para ajuste fino na estratégia de aprendizado por transferência utilizando as imagens do conjunto de treinamento. Para a etapa de validação e seleção de hiperparâmetros, adotou-se a técnica de validação cruzada estratificada *K-Fold*, com $K = 5$, garantindo que a proporção de casos displásicos e não displásicos fosse mantida em cada partição. Os hiperparâmetros de treinamento foram: 30 *epochs*, *batch size* de 16, *learning rate* de 1×10^{-5} e otimizador Adam (Kingma; Ba, 2014). Esses parâmetros foram obtidos através de *grid search*. A Tabela 1 apresenta a faixa de valores considerada para cada hiperparâmetro. Adicionalmente, foi aplicada rotação das imagens como técnica de *data augmentation* (aumento de dados). A saída da ResNet50 fornece um *score* de classificação que indica a probabilidade da presença de displasia na imagem analisada. Além disso, também gera-se um mapa de atenção que destaca as regiões da imagem que mais contribuíram para a decisão da rede a partir do algoritmo Grad-CAM.

Figura 15 – Diagrama em blocos da metodologia proposta para detecção de displasia epitelial oral utilizando o modelo base.



Fonte: do autor.

Tabela 1 – Faixa de valores dos hiperparâmetros avaliados durante o processo de *grid search* para a ResNet50.

Hiperparâmetro	Faixa de valores	Passo
Epochs	10 – 50	5
Batch size	{16, 32}	–
Learning rate	10^{-2} – 10^{-7}	10^{-1}

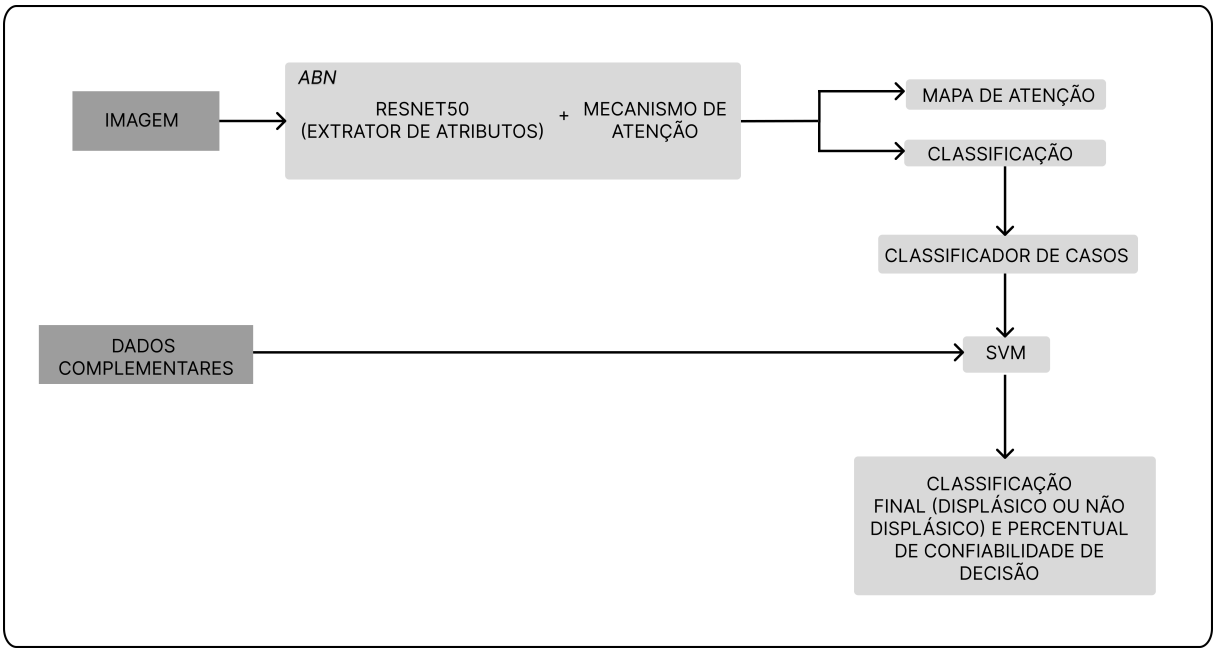
3.3.2 Modelo ABN

Além disso, outro modelo foi implementado, utilizando uma arquitetura ABN modificada com base na abordagem proposta por Mitsuhara et al. (2019). A Figura 16 apresenta o diagrama em blocos da metodologia proposta para detecção de displasia epitelial oral utilizando o modelo ABN.

O treinamento desta rede consistiu na utilização da ResNet50 do modelo base como extrator de atributos, em que a camada final de classificação foi removida, mantendo-se apenas a porção convolucional da rede. Dessa forma, a saída do extrator de atributos corresponde ao mapa de atributos gerado pela última camada convolucional, que é então processado pela ABN para gerar uma pontuação de classificação que indica a probabilidade da presença de displasia na imagem analisada, além de um mapa de atenção que destaca as regiões da imagem que mais contribuíram para a decisão da rede.

Seguindo a metodologia de Mitsuhara et al. (2019), o extrator de atributos permanece congelado durante o treinamento da ABN, ou seja, seus pesos não são atualizados. Dessa forma,

Figura 16 – Diagrama em blocos da metodologia proposta para detecção de displasia epitelial oral utilizando o modelo ABN.



Fonte: do autor.

apenas os pesos dos ramos de atenção e percepção da ABN são otimizados durante o processo de treinamento, conforme detalhado na fundamentação teórica.

Uma modificação significativa proposta por Mitsuhara et al. (2019) em relação à arquitetura ABN original e implementada neste trabalho consiste na incorporação de supervisão espacial através de mapas de atenção de referência. As anotações dos especialistas foram utilizadas para introduzir um termo adicional ($\mathscr{L}_{\text{map}}$) na função de perda do modelo, calculado a partir da comparação entre o mapa de atenção gerado pela ABN e o mapa de atenção de referência derivado das anotações do especialista. O objetivo desse termo de perda é incentivar a rede a gerar mapas de atenção mais alinhados com as regiões histopatológicas efetivamente relevantes para o diagnóstico, promovendo maior interpretabilidade e confiabilidade do modelo. Como função de perda para o mapa de atenção, foi utilizada a *Dice Loss*, métrica comumente empregada em tarefas de segmentação (Sudre et al., 2017). A nova função perda total do modelo ABN passa a ser definida conforme a Equação 3.1. Para o treinamento do modelo ABN modificado, foram utilizados os seguintes hiperparâmetros: 120 *epochs*, *batch size* de 16, *learning rate* de 1×10^{-2} e $\lambda = 0,8$ (fator de ponderação entre as perdas dos ramos de atenção e percepção). A Tabela 2 apresenta a faixa de valores considerada para cada hiperparâmetro para o modelo ABN.

$$\mathscr{L}_{\text{total}} = \lambda \mathscr{L}_{\text{attention}} + \lambda \mathscr{L}_{\text{map}} + (1-\lambda)\mathscr{L}_{\text{perception}} \tag{3.1}$$

Tabela 2 – Faixa de valores dos hiperparâmetros avaliados durante o processo de *grid search* para a ABN.

Hiperparâmetro	Faixa de valores	Passo
Epochs	60 – 180	20
Batch size	{16, 32}	–
Learning rate	10^{-2} – 10^{-4}	10^{-1}
λ	0,1 – 0,9	0,1

3.3.3 Classificador SVM

Conforme ilustrado nas Figuras 15 e 16, os dois modelos de aprendizado profundo (modelo base e modelo ABN) são complementados por um classificador SVM.

Em situações em que há mais de uma imagem por paciente, emprega-se o classificador de casos; que consiste em calcular a mediana das pontuações de classificação obtidas para cada imagem do paciente. Ou seja, dado um paciente com N imagens, realiza-se a predição da presença de displasia em cada uma delas. Em seguida, calcula-se a mediana das probabilidades atribuídas. Essa abordagem visa reduzir o impacto de possíveis erros de classificação em imagens individuais, proporcionando uma avaliação mais robusta da presença de displasia no paciente como um todo.

Em sequência, essa pontuação de classificação é combinada com os dados complementares do paciente, que são inseridos em uma SVM para obter a classificação final como displásico ou não displásico juntamente com um percentual de confiabilidade de decisão. Para o classificador SVM, utilizaram-se os seguintes hiperparâmetros: $C = 100$, *kernel* RBF e *gamma* $= 1 \times 10^{-1}$. A Tabela 3 apresenta a faixa de valores considerada para o modelo SVM. Um dos atributos utilizados no SVM é a pontuação de classificação obtida pelo modelo de aprendizado profundo. Para o treinamento, foi utilizado como pontuação o valor obtido na etapa de validação cruzada. Já para o conjunto de teste, foi utilizada a pontuação obtida a partir do modelo treinado com todo o conjunto de treinamento.

Tabela 3 – Faixa de valores dos hiperparâmetros avaliados durante o processo de *grid search* para o SVM.

Hiperparâmetro	Faixa de valores	Passo
C	{0,01 – 100}	10^1
Gamma	{0,01, 100}	10^1

3.4 ANÁLISE COMPARATIVA DOS MAPAS DE ATENÇÃO

A análise comparativa dos mapas de atenção se deu apenas para as imagens do conjunto de teste diagnosticadas como displásicas, uma vez que a região de interesse (displasia) está presente apenas nessas imagens.

Para essa etapa, utilizaram-se as máscaras binárias geradas a partir das anotações do especialista, conforme descrito anteriormente. A comparação entre os mapas de atenção gerados

pelo modelo e as máscaras binárias de referência foi realizada utilizando as métricas descritas na Seção 2.5.3, a saber: *Pointing Game*, Precisão Top K% *pixels* e Fração de Ativação.

3.5 METODOLOGIA DE IMPLEMENTAÇÃO DA PLATAFORMA EDUCACIONAL

Para viabilizar o uso educacional do modelo treinado, foi desenvolvida uma aplicação *web* interativa utilizando a biblioteca Gradio (Abid et al., 2019) da linguagem de programação Python. O modelo implementado na plataforma foi o ABN. Essa ferramenta permite que estudantes e profissionais da área da saúde realizem o carregamento (*upload*) de imagens histopatológicas, juntamente com a inserção de dados complementares, e recebam como saída a predição quanto à presença de displasia epitelial, o percentual de confiabilidade da decisão e mapas de atenção que destacam as regiões mais relevantes para a tomada de decisão do modelo, conforme ilustrado na Figura 17.

O fluxo geral de utilização da aplicação é simples e intuitivo: inicialmente, o usuário acessa a plataforma por meio de um navegador *web*; em seguida, realiza o carregamento da imagem histopatológica e fornece as informações complementares solicitadas. Após a submissão dos dados, o sistema executa automaticamente o pré-processamento da imagem, seguido da inferência utilizando o modelo previamente treinado e operando em modo de inferência. Como resultado, são exibidos simultaneamente a classificação predita, a estimativa de confiabilidade associada à decisão e os mapas de atenção gerados pela rede, permitindo ao usuário visualizar as regiões da imagem que mais contribuíram para a predição.

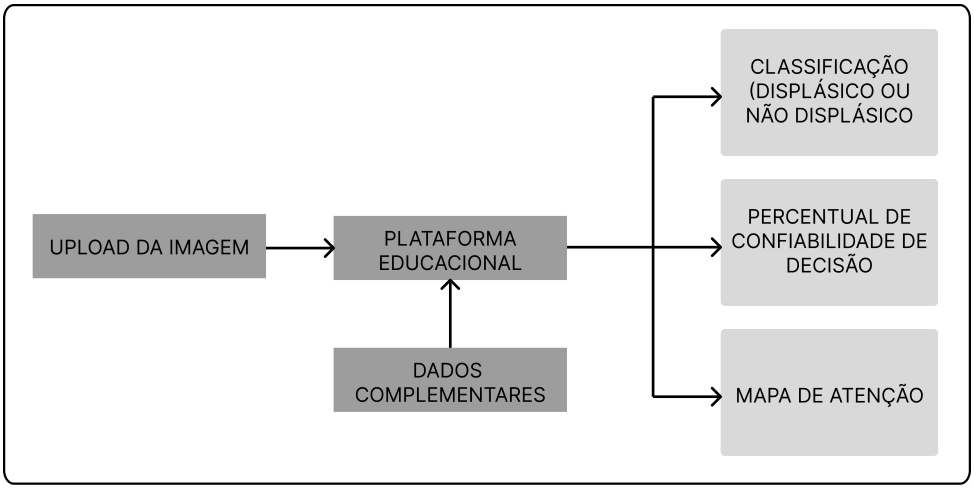
Do ponto de vista metodológico, a aplicação integra um modelo de aprendizado profundo previamente treinado e congelado, garantindo estabilidade durante o uso educacional. Todo o processamento é realizado de forma síncrona e em tempo quase real, possibilitando uma interação fluida com o usuário. A geração dos mapas de atenção é realizada sob demanda, sendo apresentada de maneira integrada à interface gráfica da aplicação, reforçando o caráter interpretável da ferramenta.

A aplicação foi disponibilizada na plataforma Hugging Face Spaces (Hugging Face, 2025), o que garante fácil acesso remoto, sem necessidade de instalação local ou configuração técnica por parte do usuário. Essa característica, aliada à interface gráfica intuitiva e ao foco em interpretabilidade, torna a ferramenta adequada tanto para uso em sala de aula quanto em ambientes de ensino a distância, funcionando como um recurso de apoio ao aprendizado e à compreensão dos padrões histopatológicos associados à displasia epitelial. O acesso à aplicação pode ser concedido mediante solicitação.

3.5.1 Metodologia de avaliação por estudantes de pós-graduação em patologia bucal

Para avaliar a utilidade das informações de apoio fornecidas pela plataforma educacional (sugestão de diagnóstico, percentual de confiança e mapa de atenção), foi conduzido um estudo com um grupo de estudantes do Programa de Pós-Graduação em Odontologia da UFSC,

Figura 17 – Esquema de utilização da plataforma educacional para detecção de displasia epitelial oral.



Fonte: do autor.

da área de concentração Estomatologia e Patologia. De maneira a simplificar a avaliação, optou-se por apresentar aos participantes apenas as informações de apoio, sem disponibilizar acesso à plataforma completa. As informações de apoio foram obtidas através do modelo ABN. O objetivo consistiu em avaliar o impacto das informações fornecidas pela ferramenta e obter retorno qualitativo acerca da utilidade dessas informações para o processo de aprendizado.

O grupo avaliador foi composto por 12 estudantes de pós-graduação *stricto sensu* (metrado e doutorado) e pós-doutorado, todos com experiência prévia em análise histopatológica de lesões orais. Após a explicação dos objetivos, riscos e benefícios do trabalho, cada estudante recebeu um Termo de Consentimento Livre e Esclarecido (TCLE) (Apêndice A), autorizando por escrito sua participação e o uso dos dados obtidos durante a execução da pesquisa.

Os participantes foram convidados a analisar um conjunto de 24 casos clínicos, sendo 12 displásicos e 12 não displásicos. Cada caso era constituído por uma imagem histopatológica acompanhada das respectivas informações clínicas. A avaliação foi conduzida de forma individual e estruturada em duas etapas.

Na primeira etapa, os participantes receberam apenas as imagens histopatológicas e as informações clínicas do caso, sendo solicitados a realizar uma classificação binária (displásico ou não displásico).

Na segunda etapa, realizada após um período de *washout* de duas semanas (Kiani et al., 2020; Bulten et al., 2021; Bellahsen-harrar et al., 2025), os participantes receberam os mesmos casos em ordem distinta, desta vez acompanhados das informações de apoio geradas pela plataforma. Foram, então, solicitados a avaliar os casos à luz das informações adicionais fornecidas. Ao término da avaliação dos casos, os participantes responderam a um questionário sobre a utilidade das informações de apoio para o suporte à tomada de decisão, considerando aspectos como: clareza da sugestão de diagnóstico, relevância do percentual de confiança e interpretabilidade dos mapas de atenção.

Para a avaliação das imagens, foi utilizado um monitor de vídeo com as seguintes

especificações: tela de 32 polegadas, resolução de 1920 $\times$ 1080 (*Full HD*), gama de cores 100% sRGB e brilho de 250 cd/m². As folhas de avaliação (disponíveis nos Apêndices B e C) foram preenchidas manualmente pelos participantes e os dados coletados foram posteriormente digitalizados e organizados para análise estatística.

As métricas de desempenho das etapas sem e com assistência foram analisadas utilizando modelos lineares de efeitos mistos com distribuição binomial, incluindo efeitos aleatórios para avaliador e caso (Steiner et al., 2020). Para avaliação de significância estatística da variável de intervenção (assistência), o valor p foi obtido através do teste de razão de verossimilhança (Steiner et al., 2018; Steiner et al., 2020).

4 RESULTADOS

Esta seção apresenta e discute os resultados obtidos a partir da metodologia descrita anteriormente. Inicialmente, são expostas as métricas de desempenho de classificação. Em seguida, são apresentadas as métricas de comparação entre os mapas de atenção gerados pelo modelo e as regiões de interesse anotadas por especialistas. Tanto o resultado de classificação quanto o de mapa de atenção são apresentados para o modelo base e para ABN. Por fim, é demonstrado um exemplo de aplicação do modelo na plataforma desenvolvida.

4.1 DESEMPENHO DE CLASSIFICAÇÃO

Os modelos foram avaliados utilizando o conjunto de teste. A partir da matriz de confusão, foram calculadas as métricas acurácia, sensibilidade, especificidade, F1-score e AUC. Para a estimativa dos intervalos de confiança de 95% e desvio padrão, empregou-se a técnica de *bootstrapping* (Rajkomar et al., 2018), que consiste em realizar amostragem com reposição a partir do conjunto de teste. Foram realizadas 5000 reamostragens para obter as estimativas dos intervalos de confiança e e os desvios padrões associados às métricas de desempenho.

A Tabela 4 apresenta os resultados obtidos para o modelo base e ABN, enquanto que a Figura 18 ilustra a distribuição das métricas de desempenho em gráficos de violino. A Figura 19 ilustra a matriz de confusão e a Figura 20 a curva ROC juntamente com a medida AUC.

Conforme apresentado na Tabela 4, os modelos alcançaram elevados valores das métricas de desempenho no conjunto de teste, com um F1-*score* médio de 0,946 para o modelo base e 0,918 para o modelo ABN. Esse último resultado se equipara ao desempenho obtido em Rocha et al. (2025) (F1-*score* de 0,91 e AUC de 0,95), sem a necessidade de segmentação manual prévia de regiões do tecido epitelial — o que pode introduzir maior erro humano no processo —, porém a um custo computacional maior.

Em uma análise preliminar, a ABN apresentou desempenho ligeiramente inferior ao modelo base, com resultados semelhantes aos reportados em Rocha et al. (2025). Contudo, o tamanho reduzido do conjunto de teste pode ter influenciado os resultados. Para avaliar se a diferença observada é estatisticamente significativa, aplicou-se o teste de McNemar, apropriado para comparar modelos pareados no mesmo conjunto de dados (Rocha et al., 2025). O teste não revelou diferença estatisticamente significativa entre os modelos, indicando desempenhos equivalentes na classificação de displasia epitelial oral. A divergência entre os modelos ocorreu em apenas um dos casos avaliados (Figura 19).

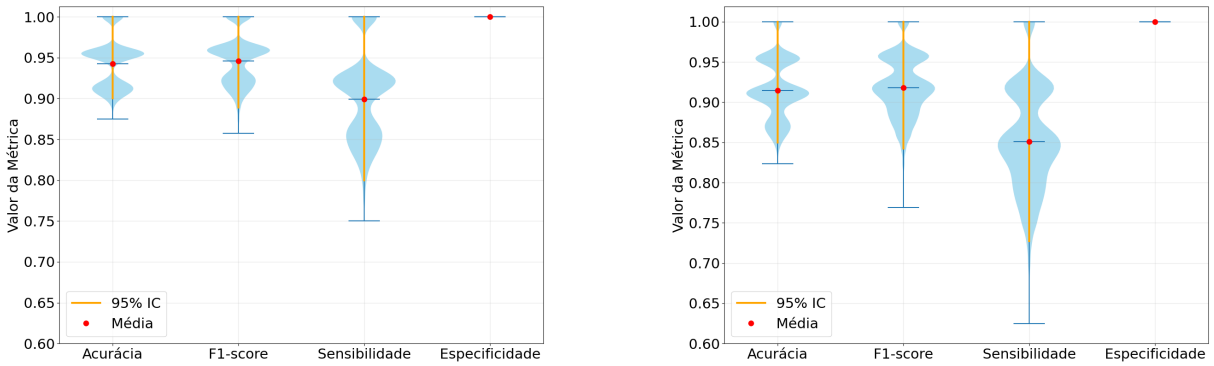
4.1.1 Influência da integração dos dados complementares

O trabalho de Rocha et al. (2025) demonstrou que a integração de dados complementares do paciente melhora o desempenho do modelo de classificação. Esse resultado também foi observado neste trabalho para os dois modelos avaliados, isto é, modelo base e ABN.

Tabela 4 – Valores médios das métricas de desempenho, desvio padrão e intervalo de confiança de 95% para o modelo base e o modelo ABN.

Métrica	Modelo Base		ABN	
	Valor médio ± DP	IC 95%	Valor médio ± DP	IC 95%
Acurácia	0,942 ± 0,030	0,900 – 1,000	0,915 ± 0,036	0,850 – 1,000
F1-score	0,946 ± 0,030	0,889 – 1,000	0,918 ± 0,037	0,842 – 1,000
Sensibilidade	0,899 ± 0,054	0,800 – 1,000	0,851 ± 0,063	0,727 – 1,000
Especificidade	1,000 ± 0,000	1,000 – 1,000	1,000 ± 0,000	1,000 – 1,000

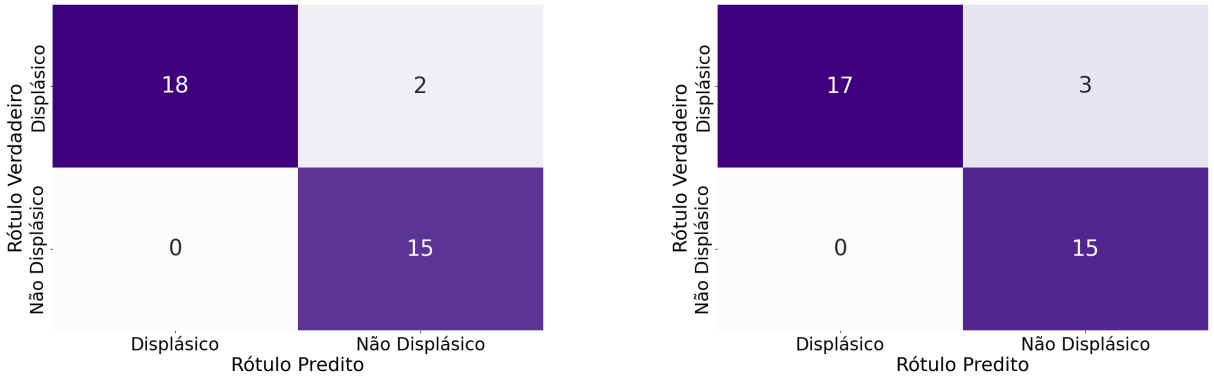
Figura 18 – Distribuição das métricas de desempenho em gráficos de violino: (a) modelo base e (b) ABN.



(a) (b)

Fonte: do autor.

Figura 19 – Matriz de confusão para: (a) modelo base e (b) ABN.



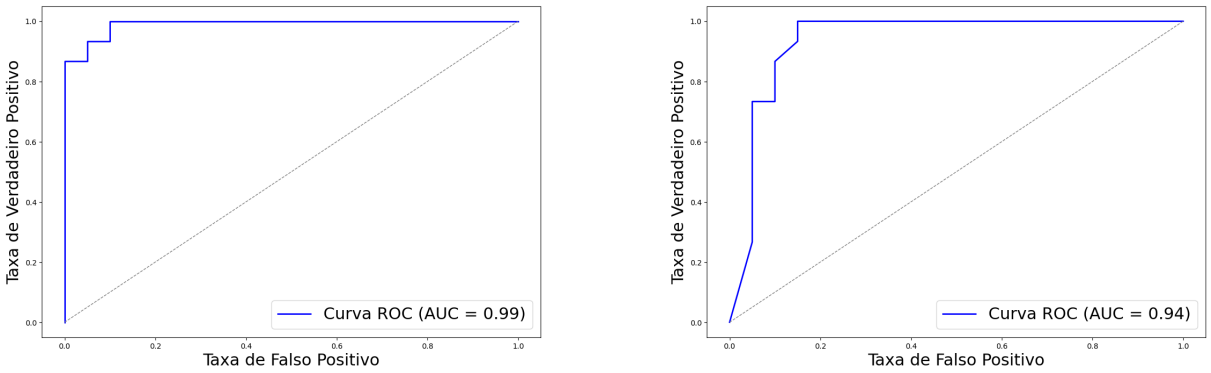
(a) (b)

Fonte: do autor.

O modelo base sem os dados complementares apresentou um F1-*score* médio de 0,89, aumentando para 0,946 ao incorporar os dados complementares. Por sua vez, modelo ABN apresentou um aumento de 0,89 para 0,918. Essa alteração se deu na mudança de falsos negativos para verdadeiros positivos, o que indica que a integração de dados complementares do paciente contribui para reduzir a ocorrência de casos de displasia epitelial não identificados pelo modelo. Para o modelo base, houve mudança de dois casos de falsos negativos para verdadeiros positivos, enquanto que para o modelo ABN houve mudança de um caso.

Em Rocha et al. (2025), o classificador SVM obteve melhor desempenho dentre outros

Figura 20 – Curva ROC e medida AUC para: (a) modelo base e (b) ABN.



(a) (b)

Fonte: do autor.

classificadores avaliados; o que serviu de motivação para escolha deste classificador neste trabalho. Entretanto, também foram avaliadas uma regressão logística e uma floresta aleatória. Os resultados obtidos para essas abordagens foram inferiores aos obtidos utilizando o SVM, com F1-*score* médio de 0,86 para a regressão logística e 0,89 para a floresta aleatória; tanto para o modelo base quanto para a ABN. Portanto, optou-se pela escolha do modelo SVM.

4.2 DESEMPENHO DOS MAPAS DE ATENÇÃO

Este trabalho traz como contribuição adicional a geração de mapas de atenção que destacam as regiões que mais influenciaram a decisão do modelo. Para avaliar a qualidade desses mapas, eles foram comparados com as anotações feitas por especialistas. É importante ressaltar que o modelo foi treinado para um procedimento de classificação e não de segmentação; de maneira que não se espera que os mapas de atenção coincidam perfeitamente com as anotações dos especialistas. O propósito dessa comparação é verificar se as regiões destacadas pelo modelo correspondem, de maneira geral, às áreas de interesse clínico e, ainda, fornecer a base para futuras melhorias.

Para o modelo base, os mapas de atenção foram obtidos através do método Grad-CAM (Selvaraju et al., 2017), que utiliza os gradientes da saída correspondente à classe prevista com relação às ativações da última camada convolucional da rede para gerar o mapa de atenção. Para o modelo ABN, os mapas de atenção são gerados diretamente pela arquitetura, conforme descrito anteriormente. Para ambos os modelos, o mapa de atenção gerado possui dimensão 17×22. De maneira a comparar os mapas gerados com as anotações dos especialistas, foi necessário redimensioná-los para o tamanho original da imagem utilizando interpolação.

Entretanto, devido à baixa resolução do mapa de atenção e o posterior redimensionamento, os métodos utilizados para obtenção dos mapas muitas vezes destacam regiões relativamente menores e incompletas de um objeto na imagem, podendo, também, falhar em detectar múltiplas ocorrências do mesmo objeto (Desai; Ramaswamy, 2020). Apesar dessas limitações,

a escolha de utilizar o Grad-CAM se justifica pela sua simplicidade de implementação, além de ser amplamente utilizado na literatura. Já para o caso da arquitetura ABN, sua escolha foi também motivada pela simplicidade de implementação, além de possibilitar um treinamento orientado às demarcações dos especialistas. Além disso, visando o contexto deste trabalho no desenvolvimento de uma ferramenta educacional, o Grad-CAM se torna a primeira opção a ser explorada, podendo ser substituído por métodos mais avançados em trabalhos futuros; enquanto que a arquitetura ABN se mostra uma abordagem inicial mais simples e com resultados promissores na literatura (Mitsuhara et al., 2019), (Song et al., 2023).

4.2.1 Modelo base

A Figura 21 ilustra os mapas de atenção gerados pelo modelo base no primeiro nível do processo de detecção de displasia epitelial; isto é, o resultado sem informação adicional de dados clínicos dos pacientes, uma vez que estes não influenciam nos mapas de atenção. Nos dois primeiros exemplos, observa-se que as regiões com maior ativação (em tons mais avermelhados) concentram-se na região limítrofe do epitélio com conjuntivo, local onde tipicamente se manifestam os primeiros sinais de displasia. No terceiro exemplo, embora o modelo não tenha identificado a displasia corretamente, a atenção foi direcionada à região epitelial, o que indica uma ativação coerente na região de interesse. Em contraste, no quarto exemplo, o modelo cometeu tanto um erro de classificação quanto de localização, concentrando-se inadequadamente na camada de queratina em vez do tecido epitelial.

De maneira a avaliar a possível falha na identificação das regiões de interesse devido às limitações mencionadas, foram adotadas duas abordagens distintas para a comparação com as anotações dos especialistas. A primeira abordagem consiste em comparar o mapa de atenção redimensionado diretamente com as anotações, enquanto a segunda abordagem envolve a aplicação de uma dilatação morfológica nas anotações antes da comparação. A dilatação morfológica é uma operação que expande as regiões anotadas, possibilitando avaliar a vizinhança das regiões de interesse. Essa técnica é especialmente útil para compensar possíveis imprecisões no mapa de atenção devido à baixa resolução e ao redimensionamento. Para esse procedimento, foi utilizado um elemento estruturante quadrado de 7×7 *pixels*. Esse parâmetro foi escolhido buscando um equilíbrio entre a expansão das regiões anotadas e a preservação da forma original, de maneira que o incremento médio da área anotada ficasse em torno de 20%, conforme ilustrado na Figura 22.

A Figura 23 apresenta os resultados da métrica de Precisão Top-K% para diferentes valores de K, comparando com as anotações originais e aquelas obtidas a partir da dilatação morfológica. No cenário original, as medianas de precisão situam-se em aproximadamente 0,38 (Top 10%), 0,31 (Top 20%) e 0,27 (Top 30%), evidenciando uma tendência decrescente esperada à medida que expandimos a região considerada. Essa diminuição ocorre porque, ao incluir porções maiores do mapa de atenção, aumenta-se também a probabilidade de incorporar regiões não relevantes. Com a dilatação das máscaras de ground truth, observa-se um aumento

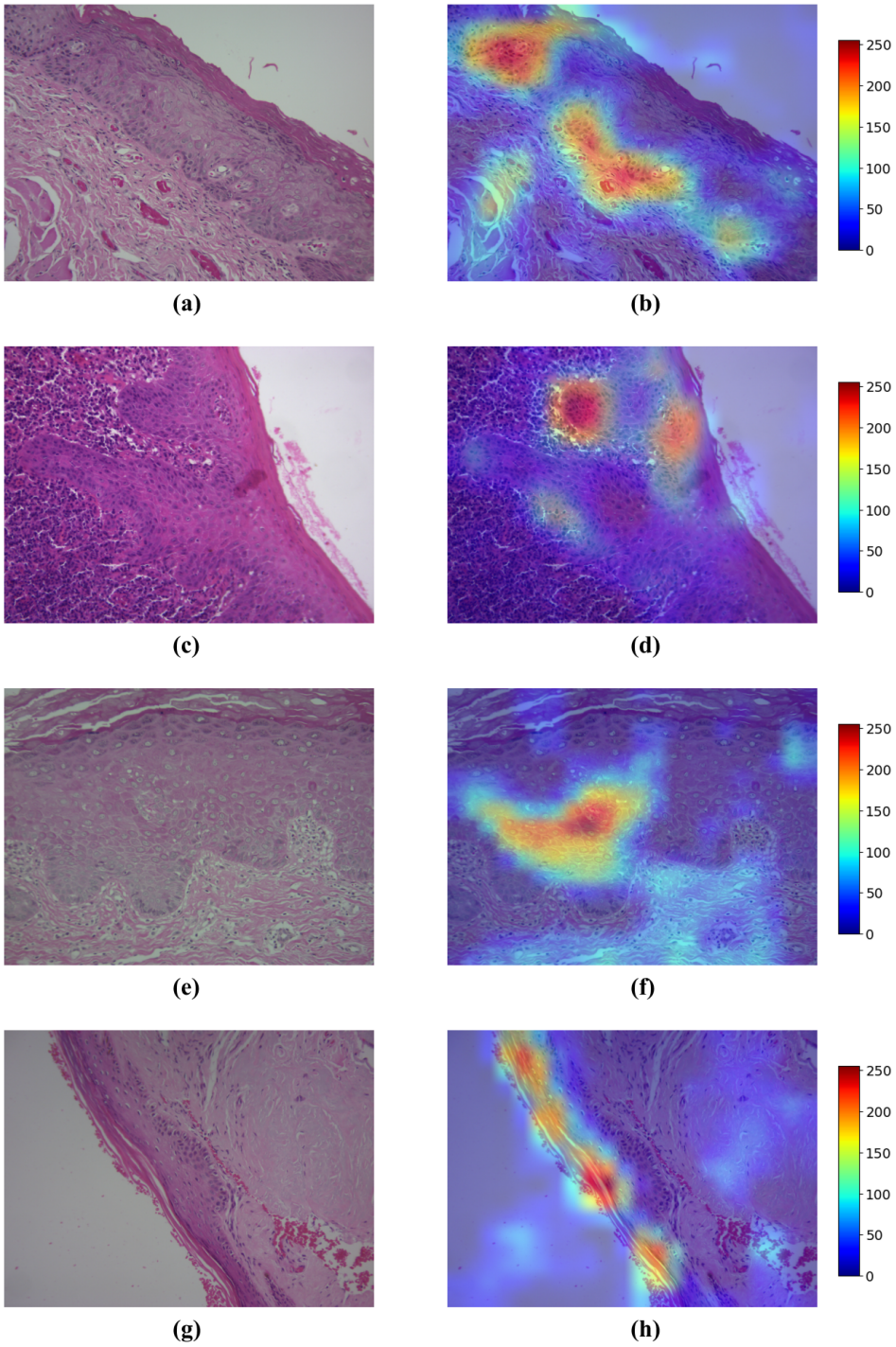


Figura 21 – Exemplos de imagem histológica e mapa de atenção provenientes do modelo base pertencentes a casos: (a-b) corretamente classificado como displásico com percentual de confiabilidade de 100%; (c-d) corretamente classificado como displásico com percentual de confiabilidade de 99%; (e-f) erroneamente classificado como não displásico com percentual de confiabilidade de 67%; (g-h) erroneamente classificado como displásico com percentual de confiabilidade de 52%.

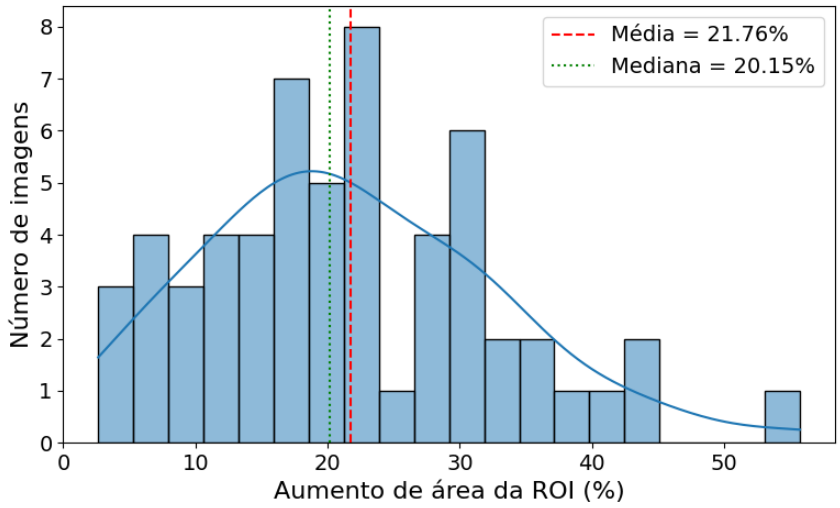


Figura 22 – Histograma de incremento percentual da área anotada após a aplicação da dilatação morfológica.

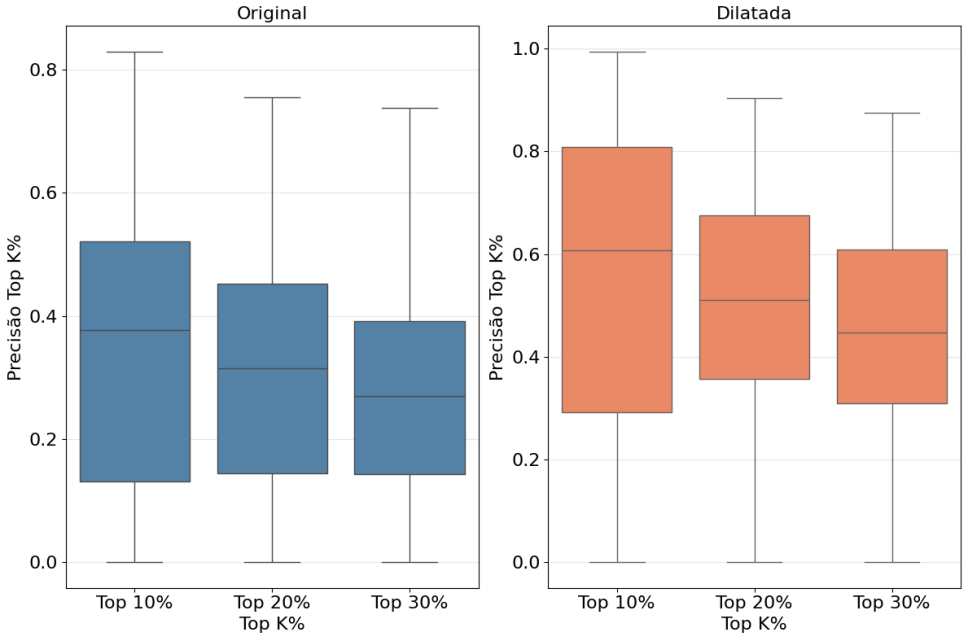


Figura 23 – Resultados da métrica de Precisão Top-K% para diferentes valores de K, para a ROI original e com dilatação para o modelo base.

expressivo nas medianas: 0,60 (Top 10%), 0,50 (Top 20%) e 0,45 (Top 30%).

Este incremento de aproximadamente 60% na precisão para o Top 10% indica que as regiões destacadas pelo Grad-CAM estão próximas das áreas de interesse, embora nem sempre coincidam exatamente com as anotações originais. Esse comportamento evidencia uma limitação do algoritmo Grad-CAM mencionada previamente, em que as regiões de ativação tendem a ser mais difusas e a cobrir áreas ligeiramente adjacentes às regiões de interesse. A dilatação, portanto, atua como uma compensação espacial, ampliando as regiões de referência e permitindo uma avaliação mais ampla da correspondência entre as atenções do modelo e as regiões relevantes da imagem.

A Figura 24 apresenta os resultados da métrica de Fração de Ativação. Observa-se que,

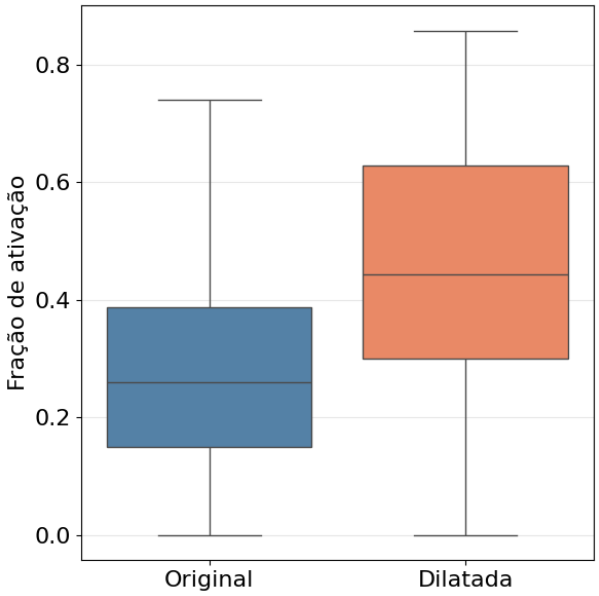


Figura 24 – Resultados da métrica de Fração de Ativação para a ROI original e com dilatação para o modelo base.

Tabela 5 – Comparação das métricas entre as máscaras binárias originais e dilatadas para o modelo base.

Métrica	Máscara binária original	Máscara binária dilatada
Mediana Precisão Top 10%	0,38	0,60
Mediana Precisão Top 20%	0,31	0,50
Mediana Precisão Top 30%	0,27	0,45
Mediana Fração de Ativação	0,35	0,55
Acurácia *Pointing Game*	0,51	0,61

no cenário original, a mediana da fração de ativação é de aproximadamente 0,35, indicando que cerca de 35% do total de concentração de ativações do modelo se encontram dentro da região de interesse. Com a dilatação das máscaras das regiões anotadas, essa mediana aumenta para cerca de 0,55. A elevação na fração de ativação após a dilatação indica que o modelo concentra uma quantidade significativa de suas ativações em áreas adjacentes às regiões de interesse, o que é consistente com os resultados observados na métrica de Precisão Top-K%.

A Figura 25 apresenta os resultados da métrica *Pointing Game*. No cenário original, a acurácia do *Pointing Game* é de aproximadamente 0,51, indicando que 51% dos picos de ativação do mapa de atenção coincidem com as regiões de interesse anotadas pelos especialistas. Após a dilatação das máscaras, essa acurácia aumenta para 0,61. Novamente, essa métrica reforça a limitação do Grad-CAM em delimitar precisamente as regiões de interesse, mas também evidencia que o modelo é capaz de identificar corretamente as áreas relevantes na maioria das vezes, especialmente quando se considera uma margem de erro espacial proporcionada pela dilatação.

A Tabela 5 resume os resultados obtidos para as três métricas avaliadas, tanto para as anotações originais quanto para as anotações dilatadas.

Os resultados obtidos — tanto em relação às anotações originais quanto às anotações

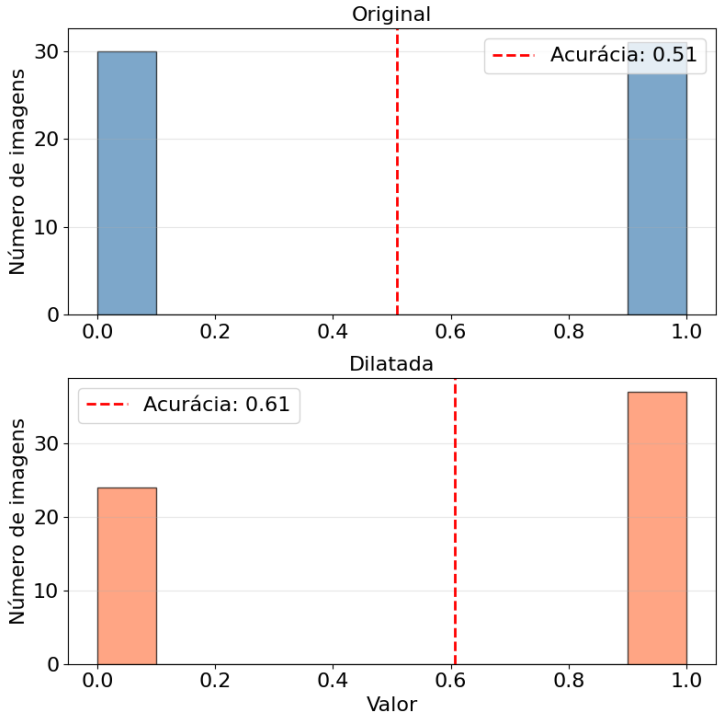


Figura 25 – Resultados da métrica *Pointing Game* para a ROI original e com dilatação para o modelo base.

dilatadas — são modestos, indicando a existência de espaço para aprimoramentos na geração dos mapas de atenção, seja por meio da aplicação de técnicas mais avançadas ou do refinamento do modelo de classificação.

4.2.2 Modelo ABN

Em se tratando do modelo base, a melhora significativa observada após a dilatação indica que, embora o Grad-CAM não delimite com precisão as fronteiras da displasia, ele identifica corretamente — na maioria das vezes — as regiões relevantes da imagem. Isso serve de motivação para a exploração de outras técnicas de geração de mapas de atenção, como a arquitetura ABN.

A Figura 26 ilustra os mapas de atenção gerados pelo modelo ABN no primeiro nível do processo de detecção de displasia epitelial. A implementação da arquitetura ABN, com um ramo de atenção dedicado, resultou em melhorias expressivas em todas as métricas avaliadas quando comparada ao modelo base utilizando Grad-CAM. A Figura 27 apresenta os resultados da métrica de Precisão Top-K% para o modelo ABN. No cenário original, as medianas de precisão situam-se em aproximadamente 0,50 (Top 10%), 0,40 (Top 20%) e 0,35 (Top 30%), representando incrementos de cerca de 32%, 29% e 30%, respectivamente, em relação ao modelo base. Com a dilatação das máscaras de *ground truth*, observa-se medianas de 0,82 (Top 10%), 0,65 (Top 20%) e 0,55 (Top 30%), o que corresponde a aumentos de 37%, 30% e 22% em relação ao cenário dilatado do modelo base.

Esses resultados indicam que o ramo de atenção permite ao modelo concentrar-se de

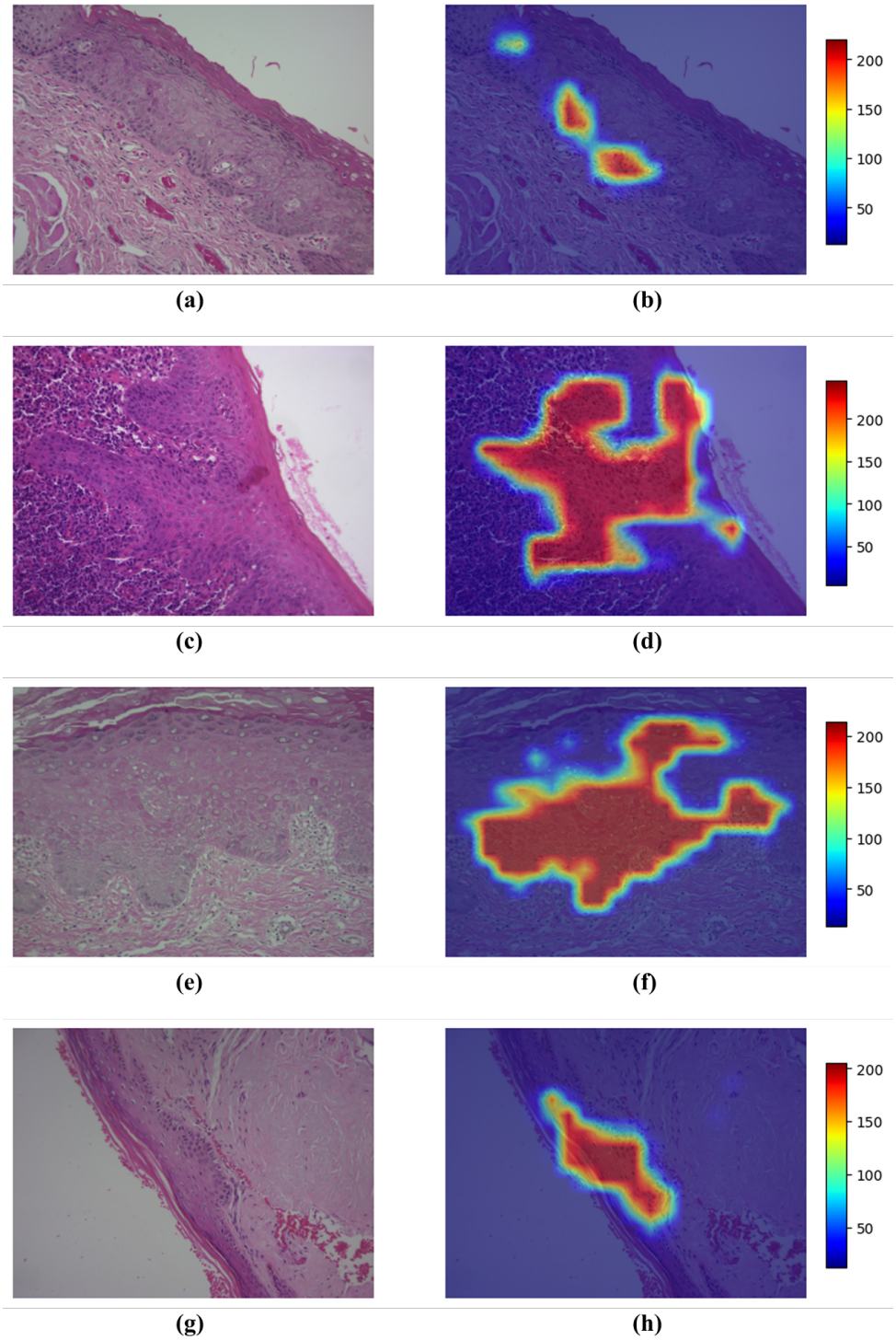


Figura 26 – Exemplos de imagem histológica e mapa de atenção provenientes do modelo ABN pertencentes a casos: (a-b) corretamente classificado como displásico com percentual de confiabilidade de 100%; (c-d) corretamente classificado como displásico com percentual de confiabilidade de 100%; (e-f) erroneamente classificado como não displásico com percentual de confiabilidade de 100%; (g-h) erroneamente classificado como não displásico com percentual de confiabilidade de 100%.

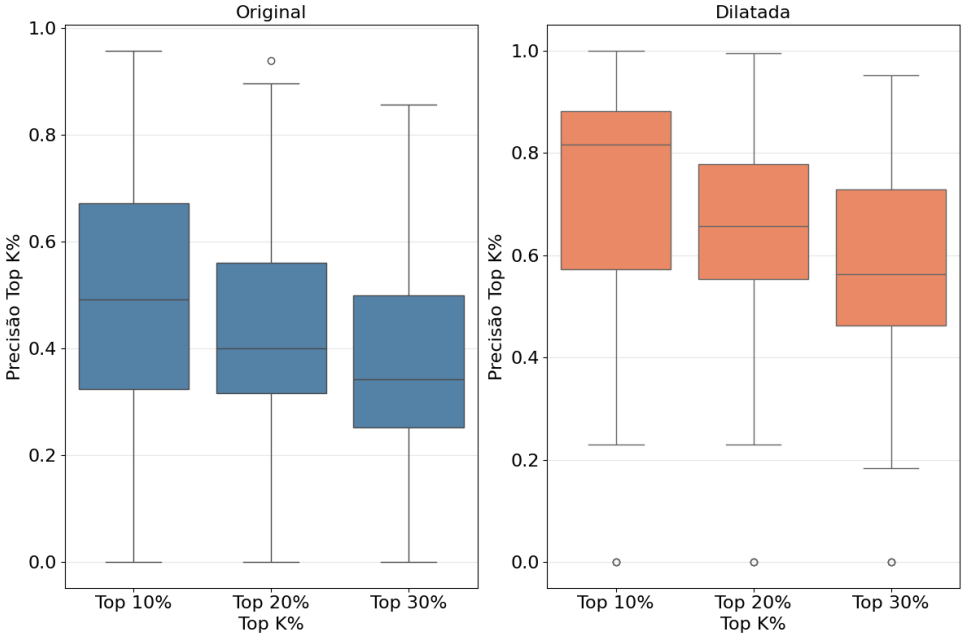


Figura 27 – Resultados da métrica de Precisão Top-K% para diferentes valores de K, para a ROI original e com dilatação para o modelo ABN.

forma mais precisa nas regiões de interesse, reduzindo a difusão das ativações observada no Grad-CAM aplicado ao modelo base. A arquitetura ABN, portanto, consegue capturar com maior fidelidade as características relevantes para a identificação da displasia epitelial.

A Figura 28 apresenta os resultados da métrica de Fração de Ativação para o modelo ABN. No cenário original, a mediana da fração de ativação é de aproximadamente 0,42, representando um aumento de 20% em relação ao modelo base (0,35). Com a dilatação das máscaras, essa mediana atinge cerca de 0,72, comparado a 0,55 no modelo base dilatado, evidenciando uma melhora de aproximadamente 31%. Esses valores demonstram que o modelo ABN não apenas identifica as regiões corretas, mas também concentra uma proporção maior de suas ativações dentro dessas áreas, reduzindo ativações espúrias em regiões irrelevantes.

A Figura 29 apresenta os resultados da métrica *Pointing Game* para o modelo ABN. No cenário original, a acurácia alcança 0,75, um incremento substancial de 47% em relação aos 0,51 obtidos pelo modelo base. Após a dilatação das máscaras, a acurácia aumenta para 0,84, superando em 38% o resultado do modelo base dilatado (0,61). Esse resultado é particularmente relevante, pois a métrica *Pointing Game* avalia a capacidade do modelo de identificar o ponto de máxima ativação dentro das regiões de interesse. A acurácia de 0,75 no cenário original indica que, em três quartos dos casos, o pico de atenção do modelo coincide precisamente com as regiões anotadas pelos especialistas.

A Tabela 6 resume os resultados obtidos para as três métricas avaliadas, tanto para as anotações originais quanto para as anotações dilatadas. A diferença ainda observada entre os cenários original e dilatado sugere que, embora o modelo ABN delimite as regiões de interesse com maior precisão que o modelo base, ainda existe espaço para aprimoramentos; seja através de ajustes arquiteturais ou exploração de funções de perda mais específicas para a tarefa de

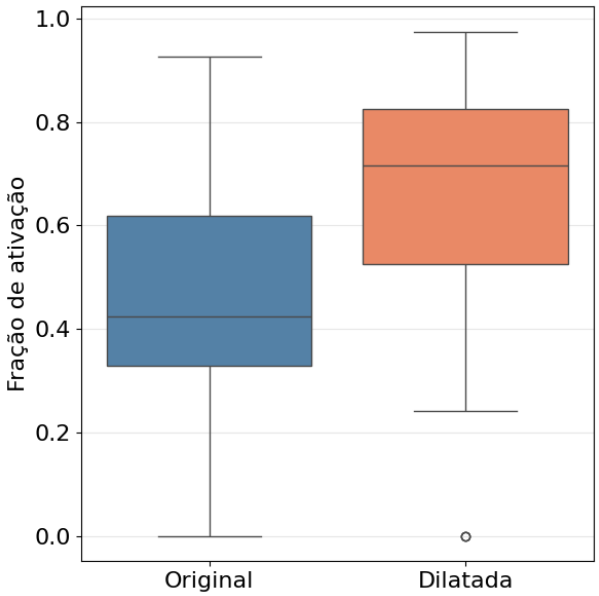


Figura 28 – Resultados da métrica de Fração de Ativação para a ROI original e com dilatação para o modelo ABN.

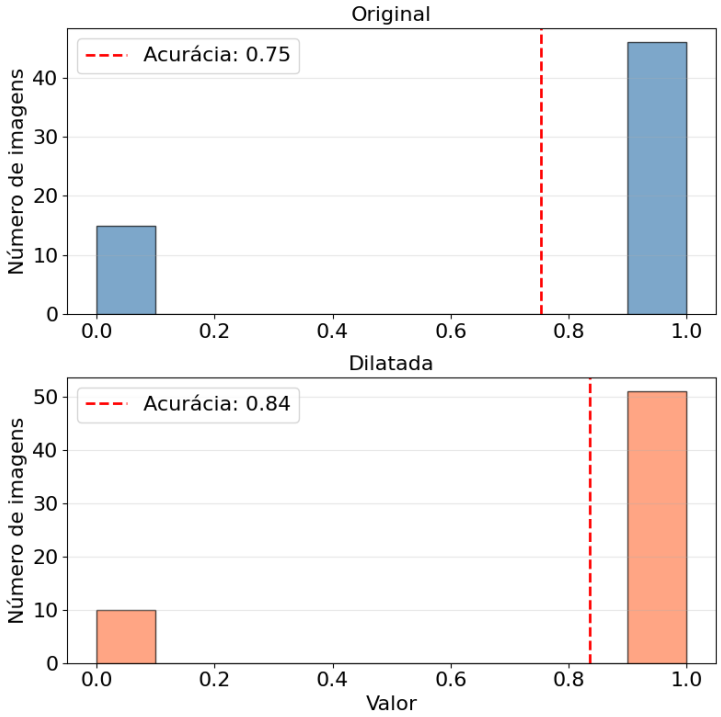


Figura 29 – Resultados da métrica *Pointing Game* para a ROI original e com dilatação para o modelo ABN.

Tabela 6 – Comparação das métricas entre as máscaras binárias originais e dilatadas para o modelo ABN.

Métrica	Máscara binária original	Máscara binária dilatada
Mediana Precisão Top 10%	0,5	0,82
Mediana Precisão Top 20%	0,4	0,65
Mediana Precisão Top 30%	0,35	0,55
Mediana Fração de Ativação	0,42	0,72
Acurácia *Pointing Game*	0,75	0,84

localização.

No entanto, os resultados obtidos já demonstram a viabilidade e eficácia da abordagem baseada em ramos de atenção dedicados para a geração de mapas de atenção para aplicações em displasia epitelial. Essa característica é essencial para a interpretabilidade clínica em contextos didáticos, pois orienta a atenção dos estudantes para as áreas de interesse sem necessariamente fornecer segmentações exatas. Além disso, é importante considerar que a demarcação precisa das regiões de displasia apresenta subjetividade e variabilidade entre especialistas, o que estabelece um limite prático para a concordância entre os mapas de atenção gerados pelo modelo e as anotações de referência. Nesse contexto, os resultados obtidos com a arquitetura ABN são promissores e indicam que os mapas de atenção gerados podem ser considerados úteis como ferramenta de apoio ao aprendizado.

4.3 PLATAFORMA EDUCACIONAL

De maneira a facilitar o aprendizado dos estudantes, a plataforma educacional foi desenvolvida para integrar o modelo de classificação e os mapas de atenção gerados, juntamente com a possibilidade de inserção adicional de dados complementares do paciente. O modelo implementado foi o ABN, em razão de apresentar melhor desempenho na geração de mapas de atenção e de apresentar desempenho de detecção de OED estatisticamente semelhante ao modelo base. O acesso à plataforma pode ser concedido mediante solicitação.

A Figura 30 apresenta uma captura de tela da interface da plataforma, destacando a visualização das imagens histopatológicas, os resultados da classificação e os mapas de atenção correspondentes. A interface foi projetada para proporcionar uma experiência intuitiva ao usuário, permitindo o carregamento de múltiplas imagens histopatológicas de um mesmo paciente através do botão "*Upload Patient Images*". Após o carregamento das imagens, o sistema processa automaticamente as informações ao clicar em "*Submit for Prediction*".

O resultado da análise é apresentado em duas seções distintas. Primeiramente, a seção "*Final Prediction Result*" exibe a classificação obtida pelo modelo acompanhada de um valor de confiabilidade, permitindo que o estudante compreenda não apenas a decisão do sistema, mas também o grau de certeza associado a ela. Logo abaixo, a plataforma oferece a opção "*Use additional patient data*", que direciona o usuário para uma interface complementar onde dados clínicos podem ser incorporados à análise.

A interface apresenta ainda duas abas de visualização: "*Original Images*" e "*Highlighted region visualizations*". A aba "*Highlighted region visualizations*" exibe os mapas de atenção (em forma de mapas de calor) sobrepostos às imagens histopatológicas originais. Esses mapas utilizam uma escala de cores que varia do azul (regiões de menor relevância) ao vermelho (regiões de maior importância para a classificação), permitindo que o estudante identifique visualmente quais áreas foram determinantes para a predição do modelo. Cada imagem processada é apresentada com seu respectivo identificador e a probabilidade de displasia, facilitando a comparação entre diferentes campos de visualização do mesmo caso.

Adicionalmente, a plataforma disponibiliza uma seção "*About*" (acessível através do botão no canto superior direito), que contém uma descrição da ferramenta, seus objetivos educacionais, desenvolvedores e instruções de uso.

A Figura 31 ilustra a seção da plataforma destinada à inserção de dados complementares do paciente para geração de uma nova predição aprimorada. Esta interface permite que o estudante incorpore informações clínicas relevantes ao caso, incluindo: idade do paciente, histórico de alcoolismo, histórico de tabagismo, localização da lesão e características clínicas da lesão. Após o preenchimento, o botão "*Make new prediction*" gera uma nova análise, cujo resultado é exibido na seção "*Enhanced Prediction Result*", apresentando a predição refinada com sua respectiva probabilidade. É importante ressaltar que a inserção desses dados complementares influencia apenas na classificação final do caso, sem impactar na geração dos mapas de atenção, que permanecem baseados exclusivamente nas características histopatológicas das imagens.

A integração do modelo de classificação, da medida de confiabilidade e dos mapas de atenção na plataforma representa um avanço significativo no processo de ensino-aprendizagem da patologia bucal. Do ponto de vista pedagógico, essa tríade de informações complementares oferece aos estudantes uma experiência de aprendizado mais profunda e reflexiva. Enquanto a predição fornece a resposta final ("displásico" ou "não displásico"), o valor de confiabilidade associado ajuda o estudante a ponderar sobre a certeza da classificação, desenvolvendo o pensamento crítico necessário para situações clínicas onde casos limítrofes são comuns. Uma probabilidade próxima de 1,0 indica alta confiança do modelo, enquanto valores intermediários sugerem casos ambíguos que merecem análise mais cuidadosa, refletindo a realidade diagnóstica em patologia oral.

A utilização efetiva da plataforma pressupõe uma abordagem pedagógica estruturada em três momentos distintos e complementares. No primeiro momento, o professor conduz a apresentação do processo diagnóstico, desde a avaliação clínica do paciente até a interpretação dos achados histopatológicos, sem o auxílio da ferramenta computacional. Essa etapa é fundamental para que o estudante construa uma base conceitual sólida e desenvolva sua capacidade de observação morfológica de forma autônoma, sem influência do sistema. No segundo momento, a plataforma é introduzida como recurso de apoio ao aprendizado: o estudante, já familiarizado com os critérios diagnósticos, tem acesso à predição do modelo, ao seu respectivo grau de confiabilidade e aos mapas de atenção, podendo confrontar sua própria interpretação com a

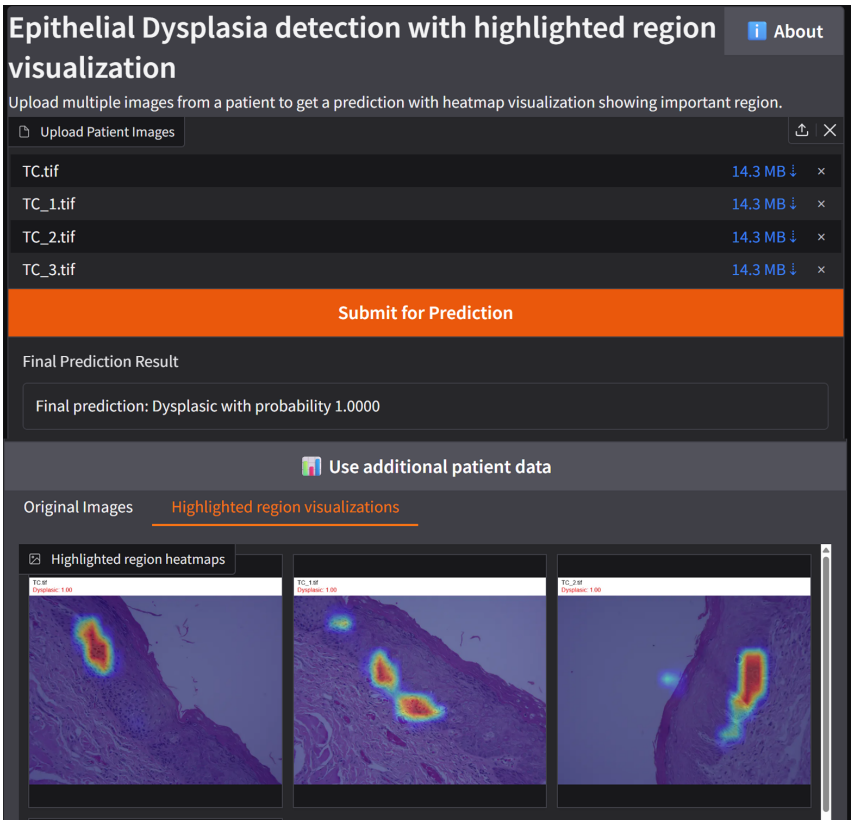


Figura 30 – Ferramenta educacional desenvolvida para auxílio na detecção de displasia epitelial oral.

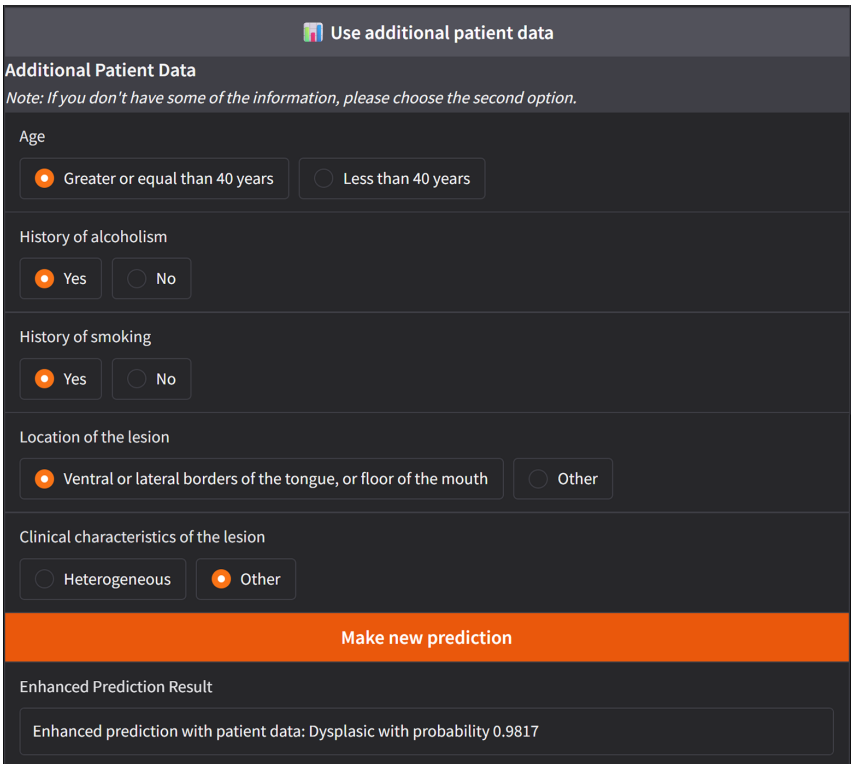


Figura 31 – Seção da plataforma para inserção de dados complementares do paciente.

resposta do sistema. Essa confrontação entre a percepção do estudante e as regiões destacadas pelo modelo estimula a reflexão crítica e a consolidação dos critérios morfológicos da displasia epitelial. Por fim, no terceiro momento — correspondente à avaliação formal da disciplina —, o estudante tem acesso exclusivamente ao caso clínico, devendo formular seu diagnóstico de forma independente, sem o suporte da ferramenta. Essa progressão metodológica, do ensino mediado pelo professor à utilização assistida pela plataforma e, finalmente, à avaliação autônoma, estrutura o aprendizado de maneira gradual.

Os mapas de atenção, por sua vez, desempenham um papel fundamental na transparência do processo decisório, permitindo que os estudantes visualizem não apenas a decisão do modelo, mas também as regiões histopatológicas que influenciaram essa decisão. Ao observar as áreas destacadas em vermelho nos mapas de calor, o estudante pode correlacionar os padrões aprendidos pelo modelo com as alterações arquiteturais e citológicas características da displasia epitelial, tais como: perda de polarização celular, pleomorfismo nuclear, aumento da relação núcleo-citoplasma e mitoses atípicas. Essa funcionalidade é crucial para o processo de aprendizado, pois facilita a compreensão dos critérios morfológicos utilizados na identificação da displasia epitelial, permitindo que o estudante desenvolva seu próprio senso crítico ao comparar sua interpretação visual com as regiões priorizadas pelo modelo.

Além disso, a plataforma promove um aprendizado ativo e autodirigido. O estudante pode testar hipóteses ao inserir diferentes combinações de dados clínicos e observar como fatores de risco conhecidos (idade avançada, tabagismo, etilismo, localização anatômica de alto risco) modificam a probabilidade final de displasia, mesmo mantendo as mesmas características histopatológicas. Essa interação reforça o conceito de que o diagnóstico em patologia oral não se baseia exclusivamente em achados microscópicos, mas sim em uma análise integrada de dados clínicos e histopatológicos. Ao experimentar com casos reais e observar variações nas predições, o estudante desenvolve habilidades de raciocínio clínico essenciais para sua futura prática profissional.

Por fim, a disponibilização dos mapas de atenção como ferramenta de apoio educacional contribui para reduzir a opacidade frequentemente associada a modelos de aprendizado profundo, transformando uma “caixa-preta” em uma ferramenta de ensino transparente e interpretável.

4.4 AVALIAÇÃO DAS INFORMAÇÕES DE APOIO DA PLATAFORMA

As métricas de desempenho dos participantes foram avaliadas, conforme descrito na Seção 3.5.1, para as duas etapas do processo de detecção de displasia epitelial: a etapa 1, onde os participantes realizaram a classificação sem acesso às informações de apoio fornecidas pela plataforma, e a etapa 2, onde os participantes tiveram acesso às informações de apoio. As métricas avaliadas incluem acurácia, sensibilidade e especificidade. Os resultados obtidos para cada métrica são apresentados nas Figuras 32, 33 e 34, respectivamente — em que cada ponto representa o desempenho de um participante — bem como resumidos na Tabela 7. Em relação

à acurácia, observa-se um aumento significativo ($p < 0,01$) na etapa 2, com uma média de 0,82 em comparação com a etapa 1, que apresentou uma média de 0,63. A especificidade também apresentou uma melhora expressiva ($p < 0,01$), passando de uma média de 0,46 na etapa 1 para 0,78 na etapa 2. Já a sensibilidade mostrou um aumento mais modesto e sem significância estatística ($p \geq 0,01$), com uma média de 0,80 na etapa 1 e 0,85 na etapa 2. Isso pode ser interpretado como uma consequência de um comportamento conservador dos patologistas orais de forma a não desprezar casos limítrofes ou duvidosos, classificando-os como displasias. O aumento significativo ($p < 0,01$) de acurácia e especificidade indica que as informações de apoio fornecidas tiveram um impacto positivo no desempenho dos participantes na detecção de OED. Além disso, a melhora expressiva na especificidade sugere que as informações de apoio foram particularmente eficazes para auxiliar os participantes na identificação de casos não displásicos. Esse resultado é de especial interesse para o sistema público de atenção à saúde, onde há uma grande demanda de casos e os recursos financeiros são restritos. O aumento da especificidade pode acarretar em significativa economia aos cofres públicos e diminuição da ansiedade aos pacientes em razão de falsos positivos.

Apesar do problema de detecção de displasia ser binário, uma avaliação estratificada por níveis de displasia pode fornecer informações adicionais sobre o desempenho dos participantes em relação a casos de diferentes graus de severidade. A displasia epitelial é comumente classificada em três níveis: leve (DEL), moderada (DEM) e severa (DES), além de casos sem displasia (SD). A avaliação estratificada por classe, vista na Tabela 8, revela que a acurácia dos participantes melhorou em todas as classes após o acesso às informações de apoio, com destaque para a classe SD, que apresentou um aumento expressivo de 0,46 para 0,78. As classes DEL, DEM e DES também apresentaram melhorias, embora em menor magnitude. Cabe ressaltar, contudo, que o número reduzido de casos por classe — consequência direta da estratificação de um conjunto já limitado de amostras — inviabiliza a realização de testes estatísticos com poder suficiente para conclusões robustas neste nível de granularidade. A obtenção de conjuntos de dados estratificados com maior número de casos por grau de severidade representa um desafio reconhecido na área e constitui uma direção relevante para estudos futuros.

Por fim, a Figura 35 apresenta os resultados do questionário de avaliação das informações de apoio, onde os participantes expressaram suas percepções sobre a utilidade e clareza das informações fornecidas pela ferramenta. De maneira geral, os participantes avaliaram positivamente as informações de apoio, destacando sua utilidade ao auxílio na detecção da OED e a clareza dos mapas de atenção, com tendências favoráveis à recomendação da utilização das informações de apoio. O principal ponto de discordância refere-se à região delimitada pelos mapas de atenção fazerem sentido do ponto de vista histopatológico e facilitarem a identificação de características histopatológicas; embora a maioria dos participantes tenha concordado com essas afirmações, houve uma parcela que expressou opiniões divergentes. Isso indica que apesar dos resultados promissores, ainda existe espaço para aprimoramento na geração dos mapas de atenção, corroborando com os resultados obtidos na avaliação quantitativa.

Uma possível limitação do estudo com os estudantes é a utilização de apenas uma ima-

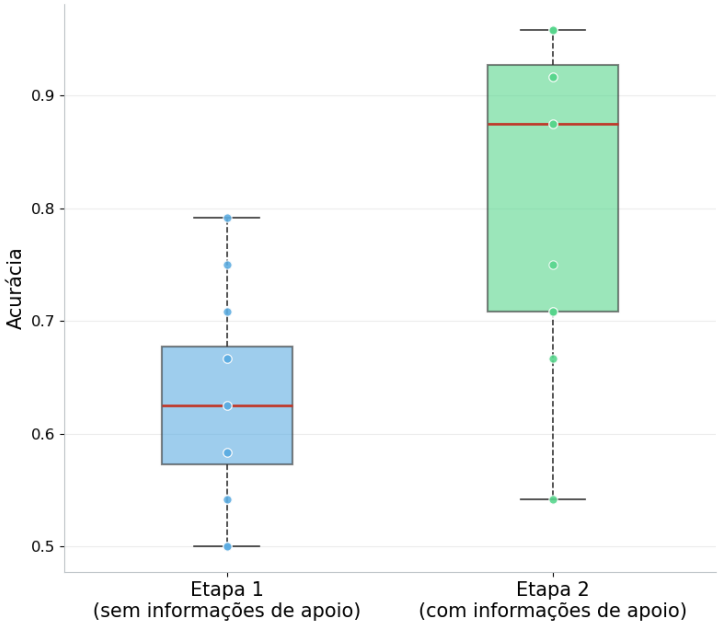


Figura 32 – Resultado da acurácia dos participantes para etapa 1 (sem informações de apoio) e etapa 2 (com informações de apoio).

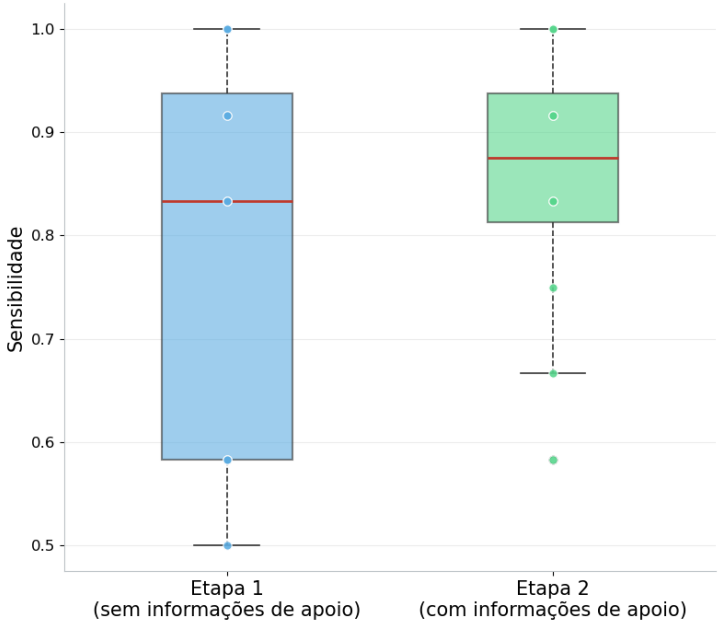


Figura 33 – Resultado da sensibilidade dos participantes para etapa 1 (sem informações de apoio) e etapa 2 (com informações de apoio).

gem histopatológica representativa por paciente. Embora essa abordagem tenha sido adotada para reduzir o tempo de avaliação, um viés relacionado à escolha da imagem representativa pode ter influenciado o desempenho dos participantes. A seleção de uma única imagem pode não capturar a heterogeneidade presente em diferentes campos de visualização do mesmo paciente, o que poderia impactar a interpretação e a classificação da displasia epitelial. Estudos futuros podem explorar a utilização de múltiplas imagens por paciente para fornecer uma visão mais abrangente e reduzir possíveis vieses na avaliação.

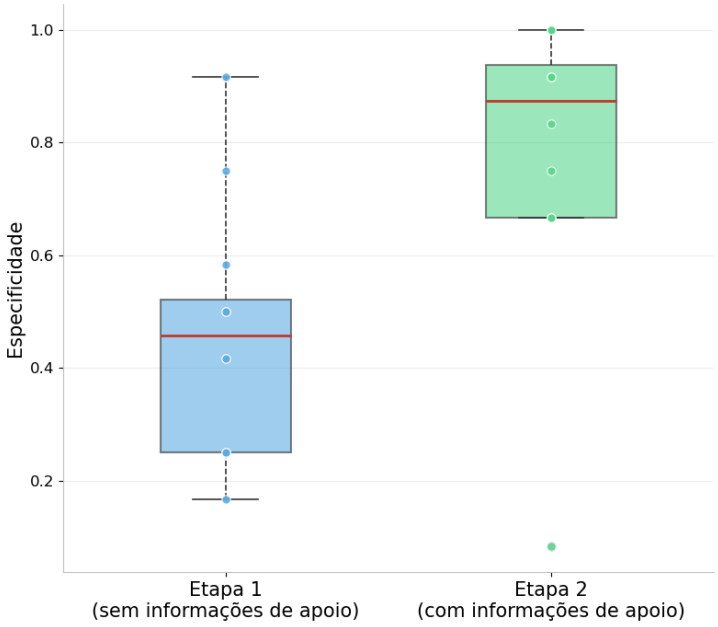


Figura 34 – Resultado da especificidade dos participantes para etapa 1 (sem informações de apoio) e etapa 2 (com informações de apoio).

Tabela 7 – Métricas de desempenho dos participantes sem e com as informações de apoio.

Métrica	Sem informações de apoio	Com informações de apoio	Valor p
Acurácia	$0{,}63 \pm 0{,}09$	$0{,}82 \pm 0{,}14$	$p < 0{,}01$
Sensibilidade	$0{,}80 \pm 0{,}19$	$0{,}85 \pm 0{,}13$	$p \geq 0{,}01$
Especificidade	$0{,}46 \pm 0{,}22$	$0{,}78 \pm 0{,}26$	$p < 0{,}01$

Tabela 8 – Acurácia dos participantes para cada classe sem e com as informações de apoio.

Classificação	Acurácia sem informação de apoio	Acurácia com informação de apoio
Displasia Epitelial Leve	0.67 ± 0.31	0.79 ± 0.18
Displasia Epitelial Moderada	0.83 ± 0.19	0.85 ± 0.23
Displasia Epitelial Severa	0.90 ± 0.17	0.92 ± 0.22
Sem Displasia	0.46 ± 0.22	0.78 ± 0.26

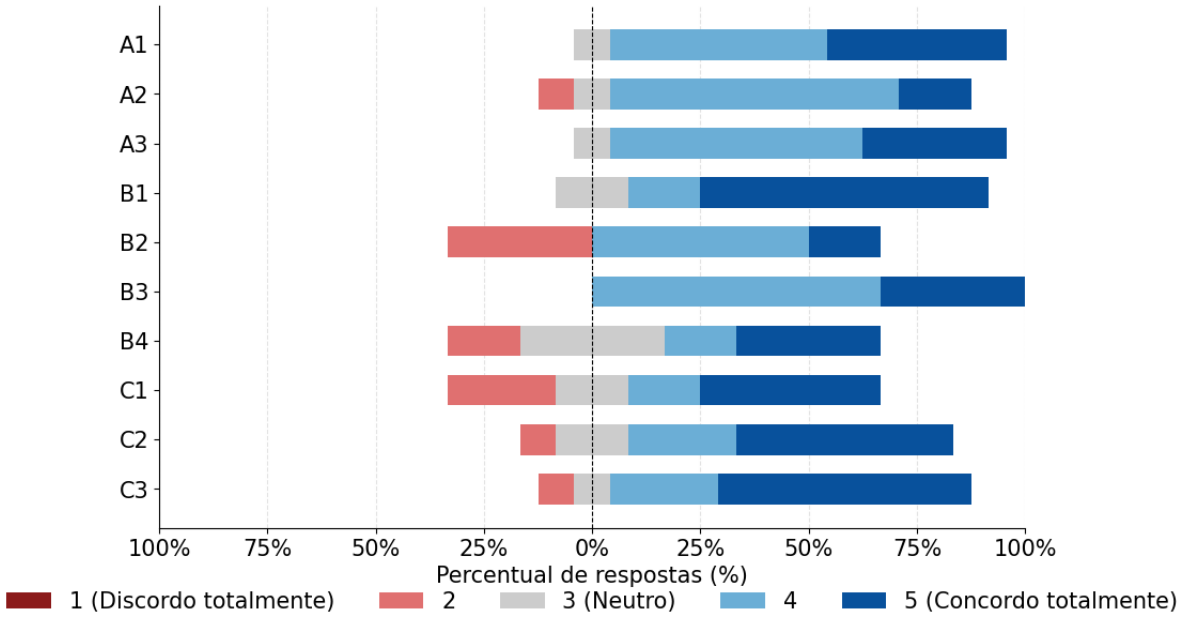


Figura 35 – Resultado do questionário de avaliação das informações de apoio.

5 CONCLUSÃO

Diante do panorama apresentado na literatura, o presente trabalho oferece contribuições significativas que endereçam lacunas identificadas em estudos anteriores. Enquanto trabalhos como os de Peng et al. (2024), Hadilou et al. (2025) e Shephard et al. (2025) apresentam mapas de atenção como recurso de interpretabilidade, nenhum deles realiza uma validação formal desses mapas através de comparação com anotações de patologistas especialistas, limitando a confiabilidade da explicabilidade proposta. Este trabalho preenche essa lacuna ao incorporar uma análise comparativa sistemática entre os mapas de atenção gerados pelo modelo e as regiões consideradas relevantes por especialistas, fortalecendo a validação da abordagem de explicabilidade adotada. Além disso, seguindo a metodologia de Rocha et al. (2025), este trabalho integra dados clínicos e imagens histológicas para detecção de displasia a nível de casos clínicos, proporcionando uma abordagem mais alinhada com a prática diagnóstica real. No entanto, enquanto Rocha et al. (2025) dependem de segmentação manual das imagens histológicas — um processo demorado e suscetível a erros —, a metodologia aqui proposta elimina essa necessidade. Ademais, em resposta à limitação de explicabilidade enfatizada na revisão de Alajaji et al. (2024), este trabalho não apenas implementa mecanismos de interpretabilidade através de mapas de atenção, mas também os integra em uma plataforma educacional interativa. Diferentemente das ferramentas desenvolvidas por Shephard et al. (2025), que focam exclusivamente na classificação e visualização, a plataforma aqui apresentada foi concebida com propósito pedagógico explícito, permitindo que estudantes interajam com predições, valores de confiabilidade e mapas de atenção de forma integrada, explorando como diferentes fatores clínicos influenciam a classificação final. Por fim, enquanto abordagens recentes com *Vision Transformers* (Hadilou et al., 2025; Shephard et al., 2025) demonstram resultados promissores, elas apresentam limitações práticas significativas, como elevado custo computacional e mapas de atenção que frequentemente destacam regiões irrelevantes (tecido conjuntivo, camada de queratina). A abordagem baseada em redes convolucionais adotada neste trabalho oferece um equilíbrio razoável entre desempenho, interpretabilidade e viabilidade de implementação em ambientes educacionais com recursos computacionais limitados.

A partir do exposto, este trabalho propôs um sistema baseado em aprendizado profundo explicável com integração de conhecimento especialista para auxiliar na detecção de displasia epitelial oral a partir de casos de leucoplasias. Este sistema integra imagens histopatológicas com dados complementares dos pacientes para auxiliar na detecção da displasia epitelial oral, além de empregar técnicas de interpretabilidade — através de mapas de calor — para fornecer informações sobre as decisões do modelo. A implementação de tais técnicas visa aumentar a confiança dos usuários na ferramenta, facilitando sua adoção em ambientes educacionais. Os resultados de classificação obtidos são equiparáveis aos de Rocha et al. (2025), com um F1-*score* de 0,92 e AUC $= 0,94$. Cabe destacar ainda que o acervo de imagens e casos aqui compilado é superior em dimensão aos bancos de dados adotados em investigações recentes de escopo semelhante (Rocha et al., 2025), os quais, por sua vez, já representavam um avanço

expressivo em termos de quantidade de imagens frente à literatura prévia. Este trabalho também contribui com a comparação quantitativa entre os mapas de atenção gerados pelo modelo e as anotações feitas por especialistas, fornecendo uma avaliação objetiva da eficácia das técnicas de interpretabilidade empregadas. Além disso, um estudo de caso indicou que a utilização das informações de apoio fornecidas pela ferramenta influenciou positivamente na classificação de OEDs realizadas por estudantes de patologia bucal, sugerindo que a ferramenta pode ser eficaz para fins educacionais.

Apesar dos avanços alcançados, os resultados obtidos pelos mapas de atenção indicam que certas vezes o modelo foca em regiões não relevantes das imagens histopatológicas, o que pode comprometer a confiabilidade das decisões. Como discutido anteriormente, essa limitação pode estar associada ao algoritmo utilizado para gerar os mapas de atenção, sugerindo a necessidade de explorar outras técnicas de interpretabilidade. Trabalhos futuros podem explorar diferentes técnicas de interpretabilidade para aprimorar a compreensão das decisões do modelo, uma vez que os resultados obtidos na detecção da displasia já são elevados. Em suma, este trabalho representa um passo importante na aplicação de aprendizado profundo para a detecção de displasia epitelial oral, destacando a importância da interpretabilidade e da usabilidade em ferramentas educacionais. As contribuições e limitações identificadas fornecem uma base sólida para futuras pesquisas e desenvolvimentos na área, com o objetivo de aprimorar a detecção precoce e o diagnóstico preciso de lesões potencialmente malignas na cavidade oral.

REFERÊNCIAS

ABDUL, N. S. et al. Applications of artificial intelligence in the field of oral and maxillofacial pathology: a systematic review and meta-analysis. **BMC Oral Health**, Springer Science and Business Media LLC, v. 24, n. 1, 2024. ISSN 1472-6831.

ABID, A. et al. Gradio: Hassle-free sharing and testing of ml models in the wild. **arXiv preprint arXiv:1906.02569**, 2019.

AFIFY, H. M.; MOHAMMED, K. K.; HASSANIEN, A. E. Novel prediction model on oscc histopathological images via deep transfer learning combined with grad-cam interpretation. **Biomedical Signal Processing and Control**, Elsevier, v. 83, p. 104704, 2023.

ALAJAJI, S. A. et al. An update on the use of artificial intelligence in digital pathology for oral epithelial dysplasia research. **Head and Neck Pathology**, Springer Science and Business Media LLC, 2024.

ARAÚJO, A. L. D. et al. Machine learning concepts applied to oral pathology and oral medicine: A convolutional neural networks' approach. **Journal of Oral Pathology & Medicine**, v. 52, n. 2, p. 109–118, 2023.

BELLAHSEN-HARRAR, Y. et al. Exploring the risks of over-reliance on ai in diagnostic pathology. what lessons can be learned to support the training of young pathologists? **PLOS One**, Public Library of Science (PLoS), v. 20, n. 8, p. e0323270, ago. 2025. ISSN 1932-6203.

BISHOP, C. M.; BISHOP, H. **Deep Learning: Foundations and Concepts**. [S.l.]: Springer, 2024.

BRUCH, J. M.; TREISTER, N. S. **Clinical Oral Medicine and Pathology**. [S.l.]: Humana Press, 2010. ISBN 9781603275200.

BULTEN, W. et al. Artificial intelligence assistance significantly improves gleason grading of prostate biopsies by pathologists. **Modern Pathology**, Elsevier BV, v. 34, n. 3, p. 660–671, 2021. ISSN 0893-3952.

CHADDAD, A. et al. Survey of explainable ai techniques in healthcare. **Sensors**, MDPI AG, v. 23, n. 2, p. 634, 2023. ISSN 1424-8220.

CHEN, L.; CHEN, P.; LIN, Z. Artificial intelligence in education: A review. **IEEE Access**, Institute of Electrical and Electronics Engineers (IEEE), v. 8, p. 75264–75278, 2020. ISSN 2169-3536.

DAWN, K. **GradCAM – Enhancing Neural Network Interpretability in the Realm of Explainable AI**. 2023. Acesso em : 07 ago. 2025. Disponível em: https://learnopencv.com/intro-to-gradcam/.

DESAI, S.; RAMASWAMY, H. G. Ablation-cam: Visual explanations for deep convolutional network via gradient-free localization. In: **2020 IEEE Winter Conference on Applications of Computer Vision (WACV)**. [S.l.: s.n.], 2020.

DWYER, B. et al. **Roboflow**. 2025. Available from https://roboflow.com. Computer vision.

FUKUI, H. et al. **Attention Branch Network: Learning of Attention Mechanism for Visual Explanation**. [S.l.]: arXiv, 2018.

GÉRON, A. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. 2. ed. [S.l.]: O'Reilly Media, 2019. ISBN 978-1492032632.

GONZALEZ, R.; WOODS, R. **Digital Image Processing, Global Edition**. [S.l.]: Pearson Education, 2018.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016.

HADILOU, M. et al. Artificial intelligence based vision transformer application for grading histopathological images of oral epithelial dysplasia: a step towards ai-driven diagnosis. **BMC Cancer**, Springer Science and Business Media LLC, v. 25, n. 1, 2025. ISSN 1471-2407.

HANKINSON, P. et al. Demystifying oral epithelial dysplasia: a histological guide. **Pathology**, Elsevier BV, v. 56, n. 1, p. 11–23, 2024. ISSN 0031-3025.

HASSELL, L. A. et al. Pathology education powered by virtual and digital transformation: Now and the future. **Archives of Pathology and Laboratory Medicine**, Archives of Pathology and Laboratory Medicine, n. 4, 2022.

HE, K. et al. **Deep Residual Learning for Image Recognition**. [S.l.]: arXiv, 2015.

Hugging Face. **Hugging Face Spaces**. 2025. https://huggingface.co/spaces. Machine Learning Applications Platform.

IOFFE, S.; SZEGEDY, C. **Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift**. [S.l.]: arXiv, 2015.

KANTHAN, R.; SENGER, J.-L. The impact of specially designed digital games-based learning in undergraduate pathology and medical education. **Archives of Pathology and Laboratory Medicine**, Archives of Pathology and Laboratory Medicine, 2011.

KHANAGAR, S. B. et al. Application and performance of artificial intelligence (ai) in oral cancer diagnosis and prediction using histopathological images: A systematic review. **Biomedicines**, MDPI AG, 2023.

KIANI, A. et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. **npj Digital Medicine**, Springer Science and Business Media LLC, v. 3, n. 1, fev. 2020. ISSN 2398-6352.

KINGMA, D. P.; BA, J. **Adam: A Method for Stochastic Optimization**. [S.l.]: arXiv, 2014.

MITSUHARA, M. et al. **Embedding Human Knowledge into Deep Neural Network via Attention Map**. [S.l.]: arXiv, 2019.

MÜLLER, D.; SOTO-REY, I.; KRAMER, F. Towards a guideline for evaluation metrics in medical image segmentation. **BMC Research Notes**, Springer Science and Business Media LLC, v. 15, n. 1, 2022. ISSN 1756-0500.

MULLER, S.; TILAKARATNE, W. M. Update from the 5th edition of the world health organization classification of head and neck tumors: Tumours of the oral cavity and mobile tongue. **Head and Neck Pathology**, Springer Science and Business Media LLC, n. 1, 2022. ISSN 1936-0568.

ODELL, E. et al. Oral epithelial dysplasia: Recognition, grading and clinical significance. **Oral Diseases**, Wiley, v. 27, n. 8, 2021. ISSN 1601-0825.

PEDROSO, C. M. et al. Machine learning models for predicting recurrence and malignant transformation of oral leukoplakia. **Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology**, Elsevier BV, v. 141, n. 2, p. 176–186, 2026. ISSN 2212-4403.

PENG, J. et al. Oral epithelial dysplasia detection and grading in oral leukoplakia using deep learning. **BMC Oral Health**, Springer Science and Business Media LLC, 2024.

PONTES, G. I. et al. Predicting malignant transformation in oral leukoplakia: A multilayer perceptron approach incorporating clinicopathological features and dna content. **Journal of Oral Pathology and Medicine**, Wiley, v. 55, n. 2, p. 249–260, 2025. ISSN 1600-0714.

RAJKOMAR, A. et al. Scalable and accurate deep learning with electronic health records. **npj Digital Medicine**, Springer Science and Business Media LLC, v. 1, 2018.

REGEZI, J. A.; SCIUBBA, J. J.; JORDAN, R. C. K. **Patologia Oral: Correlações Clinicopatológicas**. 7. ed. Rio de Janeiro: Elsevier, 2017. 402 p. ISBN 978-85-352-8705-9.

ROCHA, K. et al. A new machine-learning methodology for detection of epithelial dysplasia in oral leukoplakia lesions. **SSRN**, Elsevier BV, 2025.

ROCHA, K. da et al. A pathology-based machine learning method to assist in epithelial dysplasia diagnosis. **Research on Biomedical Engineering**, Springer Science and Business Media LLC, v. 38, n. 3, p. 989–1002, 2022. ISSN 2446-4740.

SALDIVIA-SIRACUSA, C. et al. Automated classification of oral potentially malignant disorders and oral squamous cell carcinoma using a convolutional neural network framework: a cross-sectional study. **The Lancet Regional Health - Americas**, v. 47, p. 101138–13, 2025. ISSN 2667-193X.

SELVARAJU, R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: **2017 IEEE International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2017. p. 618–626.

SHEPHARD, A. J. et al. Development and validation of an artificial intelligence-based pipeline for predicting oral epithelial dysplasia malignant transformation. **Communications Medicine**, Springer Science and Business Media LLC, v. 5, n. 1, 2025. ISSN 2730-664X.

SILVA, A. B. et al. Oralepitheliumdb: A dataset for oral epithelial dysplasia image segmentation and classification. **Journal of Imaging Informatics in Medicine**, Springer Science and Business Media LLC, 2024.

SONG, B. et al. Interpretable and reliable oral cancer classifier with attention mechanism and expert knowledge embedding via attention map. **Cancers**, MDPI AG, v. 15, n. 5, 2023. ISSN 2072-6694.

SOUZA, L. F. de. Predição da fertilidade de reprodutores suínos através de parâmetros de cinética espermática e aprendizado de máquina. Trabalho de Conclusão de Curso (Graduação em Engenharia Elétrica) – Universidade Federal de Santa Catarina. 2022.

STEINER, D. F. et al. Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. **American Journal of Surgical Pathology**, v. 42, n. 12, p. 1636–1646, 2018. ISSN 0147-5185.

STEINER, D. F. et al. Evaluation of the use of combined artificial intelligence and pathologist assessment to review and grade prostate biopsies. **JAMA Network Open**, American Medical Association (AMA), v. 3, n. 11, p. e2023267, 2020. ISSN 2574-3805.

SUDRE, C. H. et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. arXiv, 2017.

TAN, M.; LE, Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In: **Proceedings of the 36th International Conference on Machine Learning**. [S.l.]: PMLR, 2019. v. 97, p. 6105–6114.

WAAL, I. van der et al. Oral leukoplakia: a clinicopathological review. **Oral Oncology**, Elsevier BV, v. 33, n. 5, 1997. ISSN 1368-8375.

WARNAKULASURIYA, S. et al. Oral potentially malignant disorders: A consensus report from an international seminar on nomenclature and classification, convened by the who collaborating centre for oral cancer. **Oral Diseases**, Wiley, v. 27, n. 8, 2020. ISSN 1601-0825.

ZHANG, J. et al. Top-down neural attention by excitation backprop. **International Journal of Computer Vision**, Springer Science and Business Media LLC, v. 126, n. 10, 2017. ISSN 1573-1405.

ZHOU, B. et al. **Learning Deep Features for Discriminative Localization**. [S.l.]: arXiv, 2015.

APÊNDICE A – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA E ELETRÔNICA

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Nº do projeto CAAE: 80731124.7.0000.0121

Título do Projeto: Aquisição de Conhecimento Inconsciente em Análise de Imagens

Você está sendo convidado a participar de uma pesquisa cujo objetivo é a investigação da possibilidade de aquisição de conhecimento inconsciente ou não verbalizável por especialistas em análise de imagens digitais. Esta pesquisa está sendo realizada pelos estudantes Lucas Fiamoncini de Souza e Victória Carvalho Taveira da Universidade Federal de Santa Catarina, coordenada pelo professor Márcio Holsbach Costa, que se compromete a seguir a Resolução 466, de 12 de dezembro de 2012 e suas complementares.

Se você aceitar participar da pesquisa, os procedimentos envolvidos em sua participação são os seguintes: visualizar a imagem apresentada na tela do computador e prestar as informações solicitadas sobre ela de acordo com as instruções apresentadas. Ao final do processo você deverá marcar as regiões da imagem que considera mais importantes para a sua análise e responder a um questionário. Três formas poderão ser utilizadas para realizar a sua tarefa: (i) indicação na tela do computador; (ii) utilização de uma lupa virtual comandada pelo mouse; e (iii) utilização de óculos para monitoração da posição ocular. Você poderá finalizar o experimento a qualquer momento ou alterar qualquer condição que traga algum tipo de desconforto como, por exemplo, excesso de intensidade luminosa da tela do computador, desconforto físico pelo posicionamento na cadeira ou posicionamento dos óculos. O local de execução será em uma sala do tipo escritório e o tempo estimado é de trinta (30) minutos.

Os possíveis riscos ou desconfortos decorrentes da sua participação na pesquisa são desconforto físico ou mental devido ao uso dos óculos de monitoração ou por excesso de brilho da tela do computador e o cansaço devido ao tempo do experimento. De modo a minimizar ou evitar tais riscos e/ou desconfortos, os óculos podem ser retirados, as configurações de vídeo (brilho, contraste) podem ser alteradas e pausas ou a própria finalização do procedimento podem ser realizadas a qualquer momento. A

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA E ELETRÔNICA

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

quebra de sigilo, ainda que involuntária e não intencional deve ser considerada. Contudo, pretendemos garantir a manutenção do sigilo e da privacidade da sua participação durante todas as fases desta pesquisa, por meio da anonimização dos dados.

Os possíveis benefícios decorrentes da sua participação na pesquisa são a produção de conhecimento científico, a médio e longo prazo, que pode servir para aumentar o desempenho de classificadores automáticos de imagens ou o desenvolvimento de novos métodos de auxílio ao aprendizado.

Quanto ao acompanhamento e assistência, você receberá respostas e esclarecimentos a todas as perguntas e dúvidas sobre os procedimentos realizados e assuntos relacionados a essa pesquisa.

Sua participação na pesquisa é totalmente voluntária, ou seja, não é obrigatória. Caso você decida não participar, ou ainda, desistir de participar e retirar seu consentimento, não haverá nenhum prejuízo à sua relação atual ou futura com esta instituição.

Não está previsto nenhum tipo de pagamento pela sua participação na pesquisa e você não terá nenhum custo com os procedimentos envolvidos, porém, poderá ser ressarcido por despesas decorrentes de sua participação, como transporte e alimentação, cujos custos serão absorvidos pelo orçamento da pesquisa.

Caso ocorra alguma intercorrência ou dano resultante de sua participação na pesquisa, você receberá todo o atendimento necessário, sem nenhum custo pessoal. Além disso, você tem garantido o direito a solicitar indenização.

Os dados coletados durante a pesquisa serão sempre tratados confidencialmente. Os resultados serão apresentados de forma conjunta, sem a identificação dos participantes, ou seja, o seu nome não aparecerá na publicação dos resultados.

Caso você tenha dúvidas, poderá entrar em contato com o pesquisador responsável, professor **Márcio Holsbach Costa**, no Departamento de Engenharia Elétrica e Eletrônica, Centro Tecnológico, Bloco E, terceiro andar, sala 21, Universidade Federal de Santa Catarina, Campus Universitário João David Ferreira Lima, bairro Trindade, CEP 88.040-900, Florianópolis-SC, telefone **(48) 3721-2260**, e-

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA E ELETRÔNICA

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

mail **marcio.costa@ufsc.br**, ou com o **Comitê de Ética em Pesquisa com Seres Humanos** (CEPSH-UFSC), pelo telefone **(48) 3721-6094**, ou no 7º andar do Prédio Reitoria II, sala 701, de segunda à sexta, das 8h às 12h e das 14h às 18h.

Esse Termo deve ser rubricado em todas as suas páginas e assinado em duas vias, sendo uma para o participante e outra para os pesquisadores.

_____	_____
Nome do participante da pesquisa	Assinatura do participante da pesquisa
_____	_____
Nome do pesquisador	Assinatura do pesquisador
_____	_____
Nome do pesquisador responsável	Assinatura do pesquisador responsável

Local e Data: _____

APÊNDICE B – FOLHA DE AVALIAÇÃO APLICADA NA ETAPA 1 DA PESQUISA

NOME:_____ **Data:** _____

INSTRUÇÕES: Assinale sua percepção sobre a classificação do caso e seu nível de confiança associado.

CASO	CLASSIFICAÇÃO	NÍVEL DE CONFIANÇA				
		Sem confiança	Pouca confiança	Confiança moderada	Muita confiança	Totalmente confiante
1	• Não displásico • Displásico	◯	◯	◯	◯	◯
2	• Não displásico • Displásico	◯	◯	◯	◯	◯
3	• Não displásico • Displásico	◯	◯	◯	◯	◯
4	• Não displásico • Displásico	◯	◯	◯	◯	◯
5	• Não displásico • Displásico	◯	◯	◯	◯	◯
6	• Não displásico • Displásico	◯	◯	◯	◯	◯
7	• Não displásico • Displásico	◯	◯	◯	◯	◯
8	• Não displásico • Displásico	◯	◯	◯	◯	◯
9	• Não displásico • Displásico	◯	◯	◯	◯	◯
10	• Não displásico • Displásico	◯	◯	◯	◯	◯
11	• Não displásico • Displásico	◯	◯	◯	◯	◯
12	• Não displásico • Displásico	◯	◯	◯	◯	◯
13	• Não displásico • Displásico	◯	◯	◯	◯	◯
14	• Não displásico • Displásico	◯	◯	◯	◯	◯
15	• Não displásico • Displásico	◯	◯	◯	◯	◯
16	• Não displásico • Displásico	◯	◯	◯	◯	◯
17	• Não displásico • Displásico	◯	◯	◯	◯	◯
18	• Não displásico • Displásico	◯	◯	◯	◯	◯
19	• Não displásico • Displásico	◯	◯	◯	◯	◯
20	• Não displásico • Displásico	◯	◯	◯	◯	◯
21	• Não displásico • Displásico	◯	◯	◯	◯	◯
22	• Não displásico • Displásico	◯	◯	◯	◯	◯
23	• Não displásico • Displásico	◯	◯	◯	◯	◯
24	• Não displásico • Displásico	◯	◯	◯	◯	◯

APÊNDICE C – QUESTIONÁRIO APLICADO NA ETAPA 2 SOBRE AS INFORMAÇÕES DE APOIO FORNECIDAS PELO MODELO

NOME:_____ **Data:**_____

INSTRUÇÕES: Assinale sua percepção sobre as informações de apoio fornecidas pelo sistema (sugestão de diagnóstico, confiabilidade e mapa de atenção).

SEÇÃO A: UTILIDADE EDUCACIONAL	**Discordo totalmente**	**Discordo parcialmente**	**Neutro**	**Concordo parcialmente**	**Concordo totalmente**
A1. As informações de apoio (sugestão de diagnóstico, percentual de confiança e mapa de atenção) me ajudaram a identificar a presença de displasia epitelial oral.	○	○	○	○	○
A2. Os mapas de atenção facilitaram a visualização de características histopatológicas da displasia (como alterações arquiteturais e/ou citológicas).	○	○	○	○	○
A3. Eu recomendaria o uso dessas informações de apoio (sugestão de diagnóstico, percentual de confiança e mapa de atenção) para colegas que realizam análise histopatológica.	○	○	○	○	○

SEÇÃO B: INTERPRETABILIDADE DOS MAPAS DE ATENÇÃO	**Discordo totalmente**	**Discordo parcialmente**	**Neutro**	**Concordo parcialmente**	**Concordo totalmente**
B1. Os mapas de atenção (mapas de calor) são claros e fáceis de entender.	○	○	○	○	○
B2. As regiões destacadas nos mapas de atenção fazem sentido do ponto de vista histopatológico, ou seja, indicam regiões de interesse para análise histopatológica de displasias.	○	○	○	○	○
B3. Os mapas de atenção me ajudaram a entender se a sugestão de diagnóstico é relevante.	○	○	○	○	○
B4. Os mapas de atenção me ajudaram a identificar características que eu não havia notado inicialmente.	○	○	○	○	○

SEÇÃO C: CONFIABILIDADE	**Discordo totalmente**	**Discordo parcialmente**	**Neutro**	**Concordo parcialmente**	**Concordo totalmente**
C1. O percentual de confiança foi importante para que eu concordasse ou discordasse da sugestão de diagnóstico fornecida pelo sistema.	○	○	○	○	○
C2. As informações de apoio me deixaram mais confiante para analisar os casos apresentados.	○	○	○	○	○
C3. Quando foi apresentado um baixo percentual de confiança, isso me alertou para analisar o caso com mais cuidado.	○	○	○	○	○